État des lieux de la collecte et de l'exploitation des données pour la recherche et la prise de décision en santé dans les maladies rares en France

Ségolène Aymé

médecine/sciences 2018 ; 34
(hors série n° 1) : 22-25

> Nous nous accordons tous sur la nécessité d'avoir des données pour progresser dans les connaissances. Quelques aspects sont moins connus. Par exemple, 450 maladies possèdent un traitement médicalement spécifique, avec Autorisations de Mise sur le Marché (AMM) aux États-Unis ou en Europe. Il est indispensable d'évaluer l'efficacité de ces traitements en vie réelle et de les suivre. Beaucoup des produits pour les maladies rares sont innovants et coûteux. Il faut les suivre avec attention, les évaluer pour établir que c'est un bon investissement pour l'amélioration de la qualité de vie des patients. Ils représentent déjà en moyenne 4 % des dépenses pharmaceutiques, part amenée à croître, puisque nous espérons que beaucoup d'innovations arriveront sur le marché. Ce sujet est d'intérêt pour les partenariats public-privé puisque la collecte des données post-AMM est une obligation faite à l'industrie pharmaceutique et biotechnologique dans le domaine du médicament.

Les types de collections de données pour la recherche et le soin

Quelques concepts de base méritent d'être rappelés. Une base de données est une collection de données formatées et enregistrées de manière rigoureuse pour des enquêtes. Elle n'est donc pas forcément permanente. Son recrutement peut être passif, par exemple quand les données sont issues des centres de référence ou d'hôpitaux. Il peut également être actif : faute de sources de données passives suffisantes, des efforts particuliers sont nécessaires pour produire la base de données. Celle-ci peut prendre la forme d'un registre, collections transversales de données reprenant tous les patients dans une zone géographique définie, sans suivi dans le temps. Elle peut prendre la forme d'une cohorte qui se définit comme le suivi longitudinal d'une population prédéfinie sans nécessité que la définition soit géographique. La cohorte permet donc de documenter une histoire naturelle de la maladie, dans un format directement exploitable à des fins de recherche.

À chaque objectif son outil : les registres permanents ne sont pas indispensables pour atteindre des objectifs intéressants. Des enquêtes sur mesure peuvent être menées, par exemple pour établir l'incidence et la prévalence. En revanche, un registre permet de connaître ces deux variables. Il est également le meilleur levier pour assurer la surveillance,

Chargée de mission maladies rares, Institut du Cerveau et de la Moelle Épinrière, Hôpital de la Pitié-Salpêtrière, 47, boulevard de l'Hôpital, 75013 Paris, France.
segolene.ayme@icm-institute.org

mais il est très lourd à constituer. Si l'on s'intéresse à l'histoire de la maladie, les cohortes sont sans doute plus faciles à mettre en place et financer, mais les registres sont aussi de très bons outils. Des observatoires de cas sont également envisageables : les données sont recueillies au fil de l'eau, sans protocole défini ni critère. Pour autant, ces données peuvent être utiles, à défaut d'autres collections de données plus organisées.

Les différents types de registres

Plusieurs types de registres existent. Les registres de malades sont un annuaire des malades partageant le même diagnostic, par exemple les adhérents d'une association de malades. Ils sont extrêmement utiles pour recontacter les personnes souhaitant participer à des études cliniques ou répondre à des enquêtes ; ils sont le domaine de prédilection des associations de malades. Dans les registres de maladie, la base de données est constituée autour de la maladie, que le malade soit traité ou non. Cet outil est idéal pour étudier la maladie. Ces dispositifs sont généralement menés par des équipes académiques. Les registres de produits ou d'intervention sont généralement mis en place par les industriels sur demande des autorités réglementaires, qui les contraignent à fournir régulièrement des données. Un registre, pour être réellement utilisable, doit être constitué de manière aussi professionnelle qu'un essai clinique. Il faut déterminer des critères d'inclusion, des modalités de gestion des perdus de vue, être certain

de l'exhaustivité des données collectées, avoir un contrôle de qualité, avoir des moyens pour la gestion et pour l'exploitation et obéir aux contraintes réglementaires, qui ne cessent d'augmenter. Des responsabilités lourdes sont donc attachées à ces collections de données. Elles existent dès la collecte des données : il faut définir les personnes qui alimenteront la base de données. Ce sont souvent les cliniciens en contact avec les personnes malades dans les circonstances cliniques requises. Quand les moyens existent, des assistants de recherche clinique compulsent les dossiers pour compléter les données recueillies en consultation, ou vérifient et valident les données ou vont chercher, dans d'autres sources, les données manquantes. Des associations de malades peuvent également assurer ces tâches.

La gestion du registre est également une responsabilité importante : une fois la collection obtenue, il faut la faire vivre, en maintenir la qualité, assurer la sécurité des données, leur archivage et leur analyse. On considère qu'en épidémiologie, 80 % des collections de données ne sont jamais utilisées ni publiées. Cette exploitation requiert des compétences variées : il faut s'adjoindre des compétences en épidémiologie et en statistique. La responsabilité est aussi financière et administrative : il faut souvent plusieurs années d'enregistrement avant que les données soient exploitables. Or, il existe très peu de sources de financement à long terme pour ce type d'activité.

Les recommandations internationales

La collecte des données dans les maladies rares a été identifiée comme objectif dans tous les textes réglementaires et légaux. Elle figure dans la recommandation européenne de développer des registres et des bases de données, dans le texte du Conseil des Ministres de la Santé demandant à tous les pays d'avoir un plan ou une stratégie maladies rares. Le Comité d'Experts Maladies Rares de la Commission européenne a aussi publié des recommandations, en 2013, sur la façon de constituer des registres maladies rares (Figure 1). Les deux plans

français passés comportent, tous les deux, d'importants chapitres sur l'importance des données. Le premier plan avait établi un Comité national des registres maladies rares. Le deuxième avait acté la nécessité d'une banque nationale de données maladies rares et de financement des cohortes RaDiCo.

Des recommandations internationales ont été publiées récemment sur les caractéristiques que doivent respecter les collections de données pour les maladies rares, désormais adoptées par les agences de financement internationales, la Commission européenne et en France par le comité d'évaluation des registres que je préside. Les registres dans les maladies rares doivent contenir des données FAIR, « correctes » en anglais. Elles doivent être *Findable*, c'est-à-dire qu'elles doivent être publiquement accessibles. Elles doivent être *Accessible*, c'est-à-dire que les données doivent pouvoir être consultées, de même que leur compilation. Elles doivent être *Interoperable*, c'est-à-dire que des données de même nature doivent répondre au même format, qu'elles soient recueillies en France, en Finlande ou au Japon, afin de faciliter une éventuelle analyse commune. L'enjeu porte sur la compatibilité technique et sémantique, c'est-à-dire le fait que les mêmes concepts sont codés de la même manière. Les données doivent enfin être *Reusable*, c'est-à-dire qu'elles doivent pouvoir être mises à disposition d'autres personnes, d'autres groupes et d'autres chercheurs, selon des modalités explicites. Ce principe heurte les habitudes de beaucoup, qui ont consacré beaucoup d'efforts à collecter les données et ne voient pas pourquoi ils les mettraient à disposition d'un groupe concurrent. Il faut penser en termes de complémentarité : l'expérience montre que

Plus de 600 registres MR en Europe

• Couverture géographique

- Global : 10 %
- Européen : 7 %
- National : 71 %
- Régional : 12 %

• Responsabilité de la gestion

- Académique : 82 %
- Industriel : 8 %
- Privé à but non lucratif : 5 %

• Distribution par pays

AT 17 – BE 21 – BG 11 – CH 11 – DE 116 – ES 46 – FI 7 –
FR 132 – IE 11 – IT 71 – NL 25 – PL 10 – PT 16 – SE 18 – UK 86

Figure 1. Registres européens des maladies rares.

plus on met des données à la disposition d'autres groupes, plus on est associé à des exploitations de recherche, plus on publie et produit des connaissances utilisables. Ces quatre critères sont maintenant pris en compte dans l'évaluation des registres pas le Comité d'évaluation des Registres français.

Les atouts de la France

La France a la chance de posséder des registres maladies spécifiques, développés grâce à l'énergie et la détermination de beaucoup d'acteurs. Ils représentent 132 des 600 registres maladies rares en Europe mentionnés dans la brochure publiée annuellement par Orphanet. Nous avons également un grand projet de BNDMR décidé par le 2^e plan et qui parvient à maturité. Beaucoup pensent que cette base de données couvrira toutes les maladies rares et permettra la recherche clinique. Tel n'est pas le cas : cette base comprend un ensemble de données commun à l'ensemble des maladies rares, qui permettra, quand elle sera effective, une vision des patients qui se rendent dans les centres de référence et de compétences. Elle n'offrira en revanche aucune visibilité sur les patients qui ne vont pas dans ces centres, mais sont suivis par exemple par des spécialistes de ville. Elle ne comportera pas non plus de questionnaire détaillé par maladie permettant des recherches. Cette base remplit donc un objectif de santé publique et de pilotage du système de santé. Elle permettra de localiser les personnes malades dans certaines maladies spécifiques, mais ne répondra pas aux attentes de tous en matière d'amélioration des connaissances.

BaMaRa est la partie issue des données hospitalières à un stade contenant l'identité du patient, alors que les données de la BNDMR seront anonymisées et donc accessibles à des tiers pour la recherche. RaDiCo est une infrastructure de recherche issue des fonds du Grand Emprunt sous le pilotage du ministère de la Recherche. Elle s'inscrit dans les projets d'avenir présenté par la ministre, de grandes infrastructures. RADiCo aide à la collecte, à l'extraction et à la structuration des données, au contrôle de qualité, à l'optimisation de l'usage des données et à la réduction des coûts. Cette ressource a une vocation très large. Non seulement elle peut accommoder des données cliniques classiques recueillies par questionnaire, mais également des données venant d'applications mobiles, de questionnaires électroniques ou de dispositifs connectés. RaDiCo est une plate-forme prête à l'emploi et respectant tous les standards. Elle est en lien avec les ressources clés, pour faire en sorte que les données puissent être acquises à partir des centres de référence, des laboratoires de diagnostic et de recherche, des institutions de malades, de la BNDMR, d'Orphanet, etc.

RaDiCo maximise l'inter-opérabilité en respectant tous les standards internationaux. La contractualisation des relations de gouvernance de la base de données fait par exemple partie de ces standards : il faut définir qui assure la qualité des données, leur sécurité, la manière dont s'obtiennent les financements, qui est responsable des exploitations. Il faut donc une bonne gouvernance, contractualisée dans un contrat. Il faut également respecter les standards éthiques et réglementaires, mais aussi les standards techniques de sécurité des données personnelles, conformément à la nouvelle directive européenne.

L'infrastructure respecte les contraintes sémantiques et les ontologies, c'est-à-dire des vocables utilisés en science et en médecine et directement compréhensibles par les ordinateurs. RaDiCo utilise la CIM-10, mais aussi le *Human Phenome Ontology*, intégré dans le logiciel REDCap, et les codes Orpha. Cette plate-forme est déjà utilisée par les seize cohortes sélectionnées sur appel d'offres.

La situation en France n'est donc pas mauvaise, mais elle est sous-optimale. Le comité national des registres maladies rares, constitué lors du 1^{er} Plan, a par exemple été dissous lors du deuxième : les maladies rares sont rentrées dans le comité général des registres. Sur 130 registres maladies rares déclarés, seuls 12 ont fait l'objet d'une appréciation favorable par le Comité d'évaluation des Registres. Même ces derniers ne reçoivent que très peu de financement public, puisque le seul organisme finançant ces registres est l'Inserm, pour de petits montants moyens. Beaucoup d'autres collections de données n'ont jamais été évaluées, ne sont donc pas financées et leur qualité est inconnue et leur soutenabilité douteuse.

La Banque Nationale de Données Maladies Rares (BNDMR) est en développement depuis des années, ce qui a suscité une certaine frustration. Elle est à présent sur le point d'aboutir. Il reste à contractualiser avec tous les hôpitaux pour que les données remontent : on ne peut demander aux cliniciens de remplir des questionnaires supplémentaires. Un lien avec le dossier patient des centres de référence est indispensable.

RaDiCo est opérationnel, mais sa réussite est peu connue et son avenir incertain.

Des cohortes à la Fondation Imagine ont été financées par le Grand Emprunt, mais sans lien avec RaDiCo ou la BNDMR, faute de coordination au départ de tous ces projets.

Les difficultés à résoudre

La principale difficulté tient au financement : peu de sources sont disponibles. Les projets sont financés comme des projets de recherche de courte durée, de type Agence Nationale de Recherche (ANR) ou Commission européenne. Ces financements sont consacrés au développement d'outils pour chaque registre, alors que des outils génériques et des solutions existent déjà. Les budgets sont aussi consacrés à la collecte des données. Généralement, il ne reste plus de financement pour exploiter les données et faire vivre la recherche sur le long terme. Les équipes manquent de compétences épidémiologiques et d'attachés de recherche clinique (ARC) dans la plupart des lieux de collection des

données maladies rares. Il s'agit très souvent d'initiatives personnelles prises sur le temps libre des cliniciens et des chercheurs.

D'autres difficultés sont psychologiques et comportementales. La crainte du partage est très répandue : beaucoup de chercheurs voient leurs pairs comme des concurrents et se méfient des autres parties prenantes, bien que les associations de patients, les académiques et les industriels aient des intérêts convergents en la matière. Il suffit pourtant de contractualiser la gouvernance et les règles d'accès, de lecture, d'exploitation et de publication. S'asseoir sur des collections de données sans les partager représente vraiment un abus de pouvoir. Une difficulté tient à la surestimation de la valeur monétaire des données, alors que la véritable valeur tient à l'exploitation des données et à leur interprétation intelligente.

Les perspectives

S'agissant des perspectives, le règlement du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel a alourdi considérablement les obligations réglementaires des opérateurs et donc le coût des collections de données. Des dérogations s'appliquent heureusement au traitement des données de santé pour des motifs d'intérêt public, de recherche et de soins.

L'annonce du Système National des Données de Santé (SNDS) représente une véritable chance. Les maladies rares ont toute leur légitimité à entrer dans ce projet et à exploiter les données du SNDS. Ce projet vise à donner accès, au travers d'un même portail, aux données recueillies par la Caisse Nationale d'Assurance Maladie (CNAM), le Programme de Médicalisation des Systèmes d'Information (PMSI), les causes nationales de décès et les données médico-sociales. Des infrastructures sont en cours de développement notamment à l'Institut National du Cancer (INCA) et à l'Institut Nationale de la Santé et de la Recherche Médicale (Inserm) pour fournir des portes d'entrée pour tous les chercheurs. Il n'est donc pas besoin de constituer une base de données unique : les données peuvent rester là où elles sont. Il suffit de s'entendre pour que les données soient compatibles, échangeables et exploitables en commun. Finalement, les contraintes financières,

légales et réglementaires obligeront tous les acteurs à se professionnaliser, ce qui est une bonne chose.

La voie du succès réside dans une gestion multipartite des cohortes et des registres, avec une bonne gouvernance, des chartes, la contractualisation. Il faut considérer les registres comme des infrastructures et non les financer sur projet de recherche. Les registres et cohortes devraient être financés par un fonds public-privé abondé par les institutions publiques concernées, les industriels, les associations et grandes fondations.

Pour conclure, un effort de professionnalisation est nécessaire : il est temps de mettre fin aux « bricolages ». Il sera plus cher et difficile d'être en conformité, mais vous y gagnerez un meilleur soutien de vos institutions et de plus grandes facilités à nouer des partenariats avec les équipes de recherche clinique de l'hôpital, car les données seront de meilleure qualité. Il s'agit donc d'enclencher un cercle vertueux.

Je crois beaucoup dans les solutions comme celles que développe la fondation Imagine. Beaucoup de données sont à extraire de sources de données en langage naturel. Je crois également aux efforts développés par Elisa Salamanca, portant sur une exploitation optimale des données déjà existantes. Nous voyons se dessiner l'avenir. Il nous revient de faire évoluer le modèle de collections de données pour la recherche et nous assurer que nous allons vers un mode de gouvernance multipartite, pour une exploitation optimale des données grâce à des financements décents, dans un monde très contraint financièrement. ♦

State of play of French data collections in the field of rare diseases

LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.