HAL open science

La recherche clinique à partir d'entrepôts de données. L'expérience de l'Assistance Publique – Hôpitaux de Paris (AP–HP) à l'épreuve de la pandémie de Covid-19

C. Daniel

To cite this version:

C. Daniel. La recherche clinique à partir d'entrepôts de données. L'expérience de l'Assistance Publique – Hôpitaux de Paris (AP–HP) à l'épreuve de la pandémie de Covid-19. La Revue de Médecine Interne, 2020, 41, pp.303 - 307. 10.1016/j.revmed.2020.04.005 . hal-03490438

HAL Id: hal-03490438

https://hal.science/hal-03490438v1

Submitted on 20 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S024886632030120X
Manuscript_d0a160a143a148cdeb4da7d5190414ba

La recherche clinique à partir d'entrepôts de données. L'expérience de l'Assistance Publique – Hôpitaux de Paris (AP-HP) à l'épreuve de la pandémie de Covid-19.

Christel Daniel[1,2]
[1]Département Innovation et Données, AP-HP, Paris, France
[2]Sorbonne Université, Université Paris 13, Sorbonne Paris Cité, INSERM UMR_S 1142, LIMICS, Paris, France

Christel Daniel, MD, PhD
Département Innovation et Données, Direction des Systèmes d'Information
Direction – Assistance Publique – Hôpitaux de Paris
5 rue Santerre
75 012 Paris, France
Tel: +33 1 48 04 20 47
E-mail: christel.daniel@aphp.fr

© 2020 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

La recherche clinique à partir d'entrepôts de données. L'expérience de l'Assistance Publique – Hôpitaux de Paris (AP-HP) à l'épreuve de la pandémie Covid-19.

Avec la révolution numérique à l'œuvre dans le secteur de la santé, l'utilisation des données générées à l'occasion des soins réalisés en routine ouvre des perspectives nouvelles pour la recherche clinique et l'amélioration de la prise en charge des patients. Les Entrepôts de Données de Santé (EDS) qui se développent au sein des établissements de santé au plus près des patients et des professionnels de santé jouent un rôle majeur dans le développement d'un « système de santé apprenant » (Learning Health System) qui, selon l'institut de médecine américain, permet à la fois de générer régulièrement de nouvelles connaissances par l'analyse des pratiques de routine et d'intégrer continuellement les meilleures pratiques aux processus de soins [1]. A l'occasion de la pandémie Covid-19 qui a bouleversé les activités hospitalières, les systèmes d'information hospitaliers et en particulier les EDS ont dû évoluer très rapidement pour appuyer au mieux les professionnels administratifs et les soignants massivement engagés dans la réorganisation des circuits et la prise en charge des patients [2].

1 LES DONNÉES DE « VIE REELLE » : ENJEUX, AVANTAGES ET LIMITES

On désigne par données de **« vie réelle »**, des données qui ne sont pas collectées dans un cadre expérimental, mais générées à l'occasion des soins réalisés en routine et qui reflètent donc la pratique courante. De telles données peuvent provenir de sources diverses : des dossiers informatisés de patients (DPI), des bases de données médico-administratives, des registres, des cohortes, ou encore du web, des réseaux sociaux ou des objets connectés, etc. Le rapport Bégaud-Polton-von Lennep réalisé à la demande du ministère de la santé en Mai 2017 identifie les cas d'usage de données de « vie réelle » en distinguant la **recherche** – études épidémiologiques observationnelles -, l'**innovation** - production de connaissance pour développer de nouvelles modalités de prise en charge -, le **pilotage** et la régulation des soins [3].

En **recherche clinique**, contrairement à l'approche expérimentale qui repose sur la comparaison de stratégies thérapeutiques assignées à des groupes de patients par tirage au sort, l'approche observationnelle consiste à décrire et analyser un phénomène dans son cadre réel sans qu'il soit modifié par la mise en place de l'étude que ce soit la connaissance de l'hypothèse testée ou l'intervention d'un investigateur. Si les essais cliniques randomisés (ECR) restent la méthodologie de référence de la recherche biomédicale en termes de niveau

de preuve, seule une petite proportion de patients y participe et les patients recrutés sont généralement plus jeunes et ont moins de comorbidités que ceux qui ne le sont pas. Ainsi, il peut être difficile de transposer à des groupes de patients plus larges les résultats d'un ECR ou même d'intégrer le traitement ou le dispositif médical testé dans la pratique courante, limitant ainsi son utilisation malgré des preuves d'efficacité et d'innocuité apportées par les ECR [4]. Par ailleurs, les ECR sous-estiment souvent la toxicité à long terme des thérapeutiques testées, ne permettent pas d'explorer au cours du temps leur efficacité comparativement à d'autres stratégies et prennent rarement en compte les paramètres de qualité de vie [5]. Les études en « vie réelle », bénéficiant d'une plus grande généralisabilité (validité externe) sont donc de plus en plus utilisées de manière complémentaire aux ECR, notamment dans la surveillance de nouvelles thérapies (ou même de médicaments classiques) ou dans l'analyse comparative de l'efficacité de prises en charge [6]. En pharmaco-épidémiologie, des projets tels que le projet OHDSI (Observational Health Data Sciences and Informatics) démontrent l'intérêt du partage de données à large échelle dans l'analyse des usages de médicaments [7]. Si les agences du médicament (l'Agence Européenne du Médicament (EMA) et la Food and Drug Administration des États-Unis (FDA)), les académies de médecine et de nombreux travaux soulignent l'intérêt incontestable des études en « vie réelle » dans la génération de preuves en dehors du contexte des ECR [7], cela ne doit pas conduire à en ignorer les limites et notamment les différents biais susceptibles, s'ils ne sont pas pris en compte, de fausser les conclusions de la recherche. En pratique tous les biais possibles dans le contexte de l'étude (biais de sélection, d'information ou d'interprétation) doivent être discutés en présentant la méthode adoptée pour les contrôler et en cas de contrôle difficile ou impossible, une analyse de sensibilité doit permettre d'établir dans quel sens et dans quelle mesure ces biais ont pu dévier les résultats et les conclusions de l'étude [6,9]. Les recommandations de l'*European Network of Centres for Pharmacoepidemiology and Pharmacovigilance*, ENCePP (www.encepp.eu/) élaborées sous l'égide de l'EMA constituent une référence méthodologique pour la conduite d'études observationnelles.

Au-delà des études observationnelles qu'elles rendent possibles, les données de « vie réelle » représentent une opportunité pour **optimiser les essais cliniques** dont la durée et le cout ne cessent d'augmenter. Les différentes phases du cycle de vie de la recherche clinique peuvent être concernées, qu'il s'agisse de la conception du protocole, de l'analyse de la faisabilité de la recherche [10], de l'inclusion de patients [11], de la collecte de données et de la surveillance de la survenue d'évènements indésirables pendant l'essai clinque [12,13]. Bien que largement dématérialisés, les systèmes de collecte de données des essais cliniques ne se

connectent pas encore aux DPI. Ainsi les données cliniques et biologiques du DPI sont retranscrites manuellement au sein de cahiers d'observation électroniques (eCRF). La ressaisie de ces données représentant, selon le protocole de l'essai et le contenu du DPI source, 13 à 75% des données des essais cliniques, mobilise aujourd'hui de nombreuses ressources et est source d'erreurs évitables [14]. Actuellement, la plupart des DPI n'intègre pas de fonctionnalités dédiées aux activités de recherche et n'ont pas la capacité à structurer et standardiser les données pour les rendre exploitables indépendamment du contexte – soin ou recherche – de manière transparente pour l'utilisateurs [15]. C'est pourquoi, des offres commerciales de services d'optimisation des essais cliniques à destination de promoteurs de recherche clinique se développent [16, 17]. Les établissements de santé participant à ces initiatives sont amenés à identifier et standardiser les échantillons de données les plus fréquemment utilisées dans le cadre des essais cliniques ainsi qu'à en améliorer la qualité.

A l'ère du big data en santé, de l'essor des sciences de données et en particulier des nouvelles technologies d'Intelligence Artificielle (apprentissage automatique, apprentissage profond), les données de « vie réelle » sont de plus en plus utilisées afin de développer des **technologies innovantes et de nouveaux services** dans le cadre des différentes activités des professionnels de santé, qu'il s'agisse de la recherche, du soin ou de la formation. L'Intelligence Artificielle (IA) permet le développement d'aides décisionnelles diagnostique ou pronostique, notamment dans les domaines de l'imagerie, et constitue également un levier important pour le pilotage hospitalier ou l'amélioration des processus de soins. En France, la mise en place de la plateforme nationale des données de santé (GIP Health Data Hub) va être un levier puissant de développement des usages des données de « vie réelle ». Au-delà de l'ouverture du Système National des Données de Santé (SNDS) à tous les acteurs de la recherche et de l'innovation en santé, cette plateforme offrira une plus grande facilité d'appariement avec des bases de données cliniques constituées par ailleurs.

Cela dit, les données de « vie réelle » et les opportunités qu'offrent les nouvelles technologies doivent servir avant tout et de façon directe et rapide à l'amélioration continue des pratiques et à l'efficacité du service rendu aux patients. Pour progresser, les plateformes d'analyse de données doivent permettre aux communautés professionnelles de mesurer objectivement et continuellement ce qui est fait, de comparer les résultats des soins au cours du temps et entre établissements de santé, de s'interroger sur les écarts, de simuler et évaluer les impacts de certaines décisions stratégiques sur les organisations ou la qualité de la prise en charge. A ce titre, les exemples emblématiques d'utilisation de donnée de « vie réelle » décrits dans le rapport Béguard-Polton-Von Lennep [3] sont des preuves de concept de « système

apprenant rapide et réactif » (rapid learning system) : le projet CancerLinQ lancé aux Etats-Unis en 2010 par l'American Society of Clinical Oncology (ASCO) [18] et à une moindre échelle, en France, la version très aboutie du dossier communicant en cancérologie de Franche-Comté [19].

2 ENTREPOTS DE DONNÉES DE SANTÉ HOSPITALIERS : QUEL USAGE POUR LES PROFESSIONNELS DE SANTÉ ? L'EXPERIENCE DE L'ASSISTANCE PUBLIQUE – HOPITAUX DE PARIS

Avec le développement parallèle des DPI d'une part et des technologies du Big Data d'autre part, de plus en plus d'établissements de santé s'engagent dans la mise en œuvre d'**entrepôts de données de santé** (EDS) afin d'améliorer le **pilotage de l'activité hospitalière** et de faire avancer la **recherche et l'innovation** en favorisant la réalisation d'études observationnelles, la mise en place d'outils d'optimisation d'essais cliniques et le développement d'algorithmes d'aide à la décision, basés notamment sur les nouvelles technologies d'intelligence artificielle. Les EDS intègrent les informations collectées à l'occasion des prises en charges en consultation, en hospitalisation ou pour des tests diagnostiques spécialisés. Il s'agit de données très variées : données démographiques, signes vitaux, résultats d'analyses biologiques, comptes rendus et images de radiologie ou d'anatomocytopathologie, de prescriptions, dispensations ou administrations médicamenteuses, de notes ou comptes rendus cliniques de médecins ou de personnels paramédicaux, de données générés par des dispositifs médicaux (ECG, EEG, etc).

La mise en œuvre d'une infrastructure Big Data disposant des capacités de stockage et de calcul nécessaires au traitement sécurisé et performant de données de santé produites en masse et des solutions permettant de les exploiter (réalisation d'études observationnelles ou développement d'algorithme d'IA), requiert la collaboration d'experts d'un bon niveau, en nombre suffisant, et associant une pluralité de compétences : informaticiens, cliniciens, pharmacologues, épidémiologistes, statisticiens, data scientists. Un projet d'EDS représente donc un investissement très coûteux en matériel, en logiciels et en ressources humaines et impose un engagement important de la direction de l'établissement. Enfin l'importance des enjeux scientifiques, éthiques et sociétaux soulevés par les nouvelles perspectives d'exploitation de données offertes par la constitution d'un EDS nécessite la mise en place d'une gouvernance spécifique aux niveaux stratégique et opérationnel. Dans le cadre de la crise sanitaire liée à la pandémie de Covid-19, les équipes en charge des EDS au sein des

établissements de santé se sont mobilisées afin de répondre en priorité aux besoins de pilotage d'une activité hospitalière largement impactée par la mise en place en urgence de nouvelles organisations de prise en charge et la réaffectation résultante des ressources (lits, personnels, matériels, intégration de personnels volontaires, déploiement massif de la télémédecine, etc.) En parallèle, ces mêmes équipes appuyaient la mise en œuvre de projets de recherche en grand nombre initiés dans des délais réduits, sans commune mesure avec les processus habituels de la recherche clinique.

L'EDS de **l'Assistance Publique Hôpitaux de Paris (AP-HP),** premier entrepôt de données de santé hospitalier dont la constitution a été autorisée par la CNIL (Commission Nationale de l'Informatique et des Libertés) le 19 janvier 2017 à des fins de pilotage hospitalier et de recherche, rassemble aujourd'hui les données administratives et médicales de plus de 11 millions de patients [20]. La gouvernance de l'EDS de l'AP-HP associe des personnels médicaux, des chercheurs, des représentants des directions et des représentants des patients. Les règles de fonctionnement de l'EDS encadrant les conditions d'accès aux données, les modalités réalisation des projets et de valorisation des résultats des recherches ont été définies à l'issue d'une large concertation avec la communauté médicale impliquant des représentants de patients et des personnes qualifiées en éthique. Un Comité Scientifique et Ethique (CSE) évalue tous les projets de recherche multicentrique nécessitant l'exploitation de données de l'EDS. Les patients pris en charge à l'AP-HP sont informés de manière individuelle, via une mention explicite sur tous les comptes rendus hospitaliers, de l'existence et des finalités de l'EDS ainsi que des modalités d'exercice de leur droit d'opposition à l'utilisation de leurs données à des fins de recherche. Par ailleurs, un portail de transparence sur internet liste les études réalisées ou en cours de réalisation.

Les données administratives et de santé issues principalement du DPI de l'AP-HP - mais également de plusieurs dizaines d'autres applications métiers - sont progressivement intégrées au sein de l'EDS. Les données massives et volumineuses telles que les images médicales – et bientôt les données de séquençage génétique - ne sont pas dupliquées au sein de l'EDS mais mises à disposition à la demande si elles sont nécessaires à la recherche. L'intégration de ces données éparpillées au sein de différents systèmes développés en silo et non communicants, nécessite de mettre en place une chaine de structuration et de standardisation reposant sur l'adoption de modèles et de terminologies standards. Ces travaux dans le domaine de l'interopérabilité des données permettront d'intégrer des données hospitalières et de médecine de ville, les données générées au domicile des patients (données produites par le patient lui-même ou provenant d'objets connectés) et des données environnementales. Cela permettra de

disposer d'une vision plus complète du parcours de santé des patients, ce qui est un enjeu majeur pour la recherche mais avant tout pour le soin. Des solutions assurant la confidentialité des données ont été mises en place. Notamment, la désidentification ou pseudonymisation des données mise en œuvre initialement sur les données structurées a été étendue aux images médicales et aux documents textuels [21]. Des solutions d'anonymisation complète des données et de constitution de jeux de données fictives sont en cours de développement, afin d'ouvrir plus largement de manière sécurisée l'accès aux données à des fins de recherche ou d'innovation. Un des défis principaux est de trouver un juste équilibre entre confidentialité des données individuelles indispensable au respect de la vie privée des patients d'une part, et ouverture des données permettant de faire progresser la recherche pour le bien public d'autre part. Un autre enjeu majeur est l'évaluation et l'amélioration de la qualité des données, essentielle tant pour le soin que pour la recherche [22].

Les usages des solutions et services de l'EDS se développent progressivement. Le portail commun « **Pilotage** AP-HP » basé sur la solution Cognos (IBM) déjà utilisée pour le pilotage stratégique et médico-économique a été mis à jour et intégré à la plateforme Big Data pour offrir à des utilisateurs plus nombreux se situant au plus près des professionnels de santé des fonctionnalités de suivi de l'activité et des performances sur un périmètre de données plus large. Dans le domaine de la **recherche**, en trois ans, plus de 70 demandes d'accès aux données de l'EDS pour des projets de recherche ou d'innovation ont été traitées par le CSE. Les recherches en cours sont principalement des **études épidémiologiques** de différents types : études d'incidence, analyse de comorbidités, études de corrélation entre facteurs de risque et survenue de maladies ou de complications, études de pratiques. Dans le **domaine de l'Intelligence Artificielle (IA),** les pistes en exploration pour le développement d'aides décisionnelles sont nombreuses, notamment dans les domaines de la réanimation, de l'imagerie et de l'amélioration des processus de soins. Un projet en cours porte, par exemple, sur la prédiction de risque de fracture d'origine ostéoporotique grâce à un algorithme permettant un calcul automatique de la densité osseuse sur des scanners abdominaux. Des équipes de recherche travaillent par ailleurs sur des algorithmes permettant de mieux dépister les cancers en exploitant soit des données d'imagerie (mammographies, tomosynthèses, IRM), soit des données d'analyses sanguines.

Les données de l'EDS sont également utilisées pour **optimiser l'organisation des essais cliniques.** Dans le cadre du projet EHR2EDC (https://eithealth.eu/project/ehr2edc), l'AP-HP a participé à une vaste étude d'évaluation d'une solution de connectivité entre DPI et bases de données d'essais cliniques, testée dans le cadre de six essais cliniques multicentriques

conduits par trois industriels pharmaceutique dans quatre hôpitaux distincts, démontrant que plus de 15% des données des essais pouvaient être collectées de manière semi-automatique à partir du DPI.

Un mois après l'admission à l'AP-HP du premier patient diagnostiqué Covid-19, une base de données spécifique – EDS-COVID - a été constituée à partir de l'EDS. En complément des données démographiques, de prise en charge (parcours du patient), du PMSI (diagnostics et actes), de biologie, de prescription médicamenteuse, la base de données EDS-COVID intègre des données d'intérêt automatiquement extraites des comptes rendus en utilisant des techniques de traitement automatique du langage, ainsi que certaines images médicales, en particuliers les scanners thoraciques des patients porteurs ou suspectés de Covid-19. D'autres sources de données sont intégrées progressivement telles que la base de données Covidom (application de télé suivi à domicile des patients sans signe de gravité) ou les bases de données de gestion des ressources humaines, des médicaments et dispositifs médicaux. Une mention d'information spécifique à la base EDS-COVID a été ajoutée au site de l'EDS (https://eds.aphp.fr). Les données de la base EDS-COVID alimentent le portail commun « Pilotage AP-HP » et ont permis le développement de tableaux de bords de pilotage informant quotidiennement la cellule de crise sur les caractéristiques démographiques, cliniques et biologiques des patients diagnostiqués Covid-19, l'évolution de leur état de santé, les modalités de leur prise en charge et l'évolution des capacités d'accueil et de traitement au sein de l'hôpital. Dans le domaine de la recherche, plus de vingt demandes d'accès aux données de la base EDS-COVID pour des projets de recherche ont été traitées par le CSE selon une procédure accélérée. Depuis sa mise en place à l'AP-HP dès le début de la crise, le Comité de pilotage recherche Covid-19 a par ailleurs priorisé 14 essais thérapeutiques et 23 autres études. Une solution de transfert de données est en cours de développement afin de permettre l'alimentation à partir de la base EDS-COVID des bases de données de ces essais cliniques ou cohortes. Enfin la standardisation des données de la base EDS-COVID selon des standards internationaux (modèle OMOP, terminologies CIM10, LOINC, ATC, etc) a permis à l'AP-HP d'intégrer le consortium international 4CE (Consortium for Clinical Characterization of Covid19 by EHR (4CE), https://transmartfoundation.org/covid-19-community-project) dont l'objectif est de partager et analyser rapidement à grande échelle des données agrégées de patients hospitalisés diagnostiqués Covid-19 afin de mieux caractériser cette maladie.

Le tableau 1 résume les modalités d'accès aux données de l'EDS AP-HP et la manière dont elles ont été aménagées dans le cadre de la crise sanitaire liée à la pandémie COVID19.

Au-delà de ces usages locaux et des bénéfices directs apportés aux professionnels de santé de l'institution, l'EDS de l'AP-HP, au même titre que d'autres EDS hospitaliers, a vocation à constituer un composant majeur de plateformes de partage et d'exploitation de données à plus grande échelle. Dans le cadre de cinq projets pilotes de la plateforme des données nationale (Health Data Hub), des échantillons de données sont construits et fiabilisés au sein de l'EDS de l'AP-HP avant leur mise à disposition sur un serveur sécurisé au sein des locaux de la DREES. Dans le contexte international du projet EHR2EDC, un large échantillon de données standardisé en lien avec des partenaires Européens et Nord-Américains doit permettre de réaliser des études de faisabilité d'essais cliniques et d'augmenter le nombre de patients y participant. En tant que « nœuds » de tels systèmes nationaux ou internationaux, les EDS hospitaliers en sont les partenaires essentiels, à plusieurs titres. En tant que « producteurs de données », ils ont la charge de collecter, standardiser et fiabiliser des données cliniques partagées en garantissant aux patients la sécurité des traitements et le bon usage de données dont ils sont dépositaires. En tant que « promoteurs de recherche », ils sont le plus souvent associés aux recherches réalisées sur les données qu'ils partagent, apportant au sein d'équipes pluridisciplinaires de recherche non seulement leur expertise médicale, scientifique et méthodologique mais également leur connaissance des données nécessaire au pré-traitement, à l'enrichissement d'échantillons de données pour la recherche et l'innovation, et surtout à l'identification et au contrôle de biais susceptibles d'avoir un impact sur le résultat de la recherche ou la qualité des algorithmes développés.

3 CONCLUSION

L'utilisation des données de vie réelle, qu'il s'agisse de réaliser des études observationnelles ou de développer des aides décisionnelles diagnostiques ou pronostiques, va se multiplier et prendre une place croissante : c'est un mouvement déjà engagé au plan international, et qui va s'accroître avec les possibilités offertes par les technologies numériques et l'émergence de plateformes – institutionnelles, académiques ou commerciales - de partage de données à grande échelle.

L'EDS de l'AP-HP s'est constitué d'emblée dans la perspective d'un « système apprenant rapide et réactif », avec l'objectif de mettre à disposition des professionnels des hôpitaux des solutions d'accès et d'analyse de l'ensemble des données de santé consolidées et enrichies,

quel que soit le contexte – soin ou recherche – de leur collecte. Ces solutions de visualisation de données à l'échelle individuelle ou populationnelle, d'identification de patients similaires, d'aides décisionnelles, etc. - doivent leur permettre de personnaliser et d'améliorer la prise en charge des patients. Dans le contexte de la pandémie, l'équipe en charge de l'EDS a été amenée à consolider une base de données dédiée au Covid-19 en moins d'un mois, à l'enrichir par près de vingt sources de données complémentaires et à développer de nouvelles solutions de pilotage de l'activité hospitalière à partir de ses données. L'appui de personnels des sites hospitaliers et de volontaires externes a été déterminant. La base de données EDS COVID va également permettre la réalisation d'études observationnelles AP-HP ou multicentriques nationales et internationales et l'appui à des essais thérapeutiques, toujours dans le cadre de la gouvernance existante et sans déroger aux règles d'utilisation des données, en particulier à l'avis favorable du CSE. La crise sanitaire actuelle met en évidence la nécessité pour les établissements de santé de poursuivre la mise en place d'EDS et de renforcer leur expertise en science des données. En effet, investir dans des plateformes d'analyse, au plus proche des professionnels de santé, à l'instar d'organisations reconnues internationalement pour l'excellence des services qu'elles offrent (comme la Mayo Clinic, Kaiser Permanente, Intermountain Health Care aux Etats-Unis), leur permettra de s'évaluer, de s'améliorer et d'inventer la médecine de demain.

Remerciements :

L'auteure remercie l'équipe de l'Entrepôt de Données de Santé de l'AP-HP et l'ensemble des personnels de l'AP-HP et des volontaires ayant contribué à la mise en œuvre de la base de données EDS-COVID et des solutions d'exploitation de cette base de données.

Conflits d'intérêt :

Aucun

Références :

1. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. J Am Med Inform Assoc. 2015;22(1):43-50.
2. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M et al. Rapid Response to COVID-19: Health Informatics Support for Outbreak Management in an Academic Health System. J Am Med Inform Assoc. 2020 Mar 24. pii: ocaa037. doi: 10.1093/jamia/ocaa037. [Epub ahead of print].
3. Bégaud B, Polton D, von Lennep F. Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé. Rapport réalisé à la demande de Madame la Ministre de la santé Marisol Touraine, remis en Mai 2017. https://solidarites-sante.gouv.fr/IMG/pdf/rapport_donnees_de_vie_reelle_medicaments_mai_2017vf.pdf
4. Frieden TR Evidence for health decision making—beyond randomized, controlled trials. N Engl J Med. 2017; 377:465–475
5. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br JCancer. 2014; 110:551–555
6. Visvanathan K, Levit LA, Raghavan D et al. Untapped potential of observational research to Inform clinical decision making: American Society of Clinical Oncology research statement. J Clin Oncol. 2017; 35(16):1845–1854
7. You SC, Jung S, Swerdel JN, Ryan PB, Schuemie MJ, Suchard MA et al. Comparison of First-Line Dual Combination Treatments in Hypertension: Real-World Evidence from Multinational Heterogeneous Cohorts. Korean Circ J. 2020;50(1):52-68.
8. Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data: Motivation-challenges-success factors. Onkologe (Berl). 2018;24(Suppl 2):91-98.
9. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE special task force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017;26:1033–1039
10. Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med. 2009;48(1):38-44.
11. Schreiweis B, Trinczek B, Köpcke F, Leusch T, Majeed RW, Wenk J, et al. Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials. Int J Med Inf.2014;83(11):860-8.
12. Kush R, Alschuler L, Ruggeri R, Cassells S, Gupta N, Bain L, et al. Implementing Single Source: the STARBRITE proof-of-concept study. J Am Med Inform Assoc JAMIA. 2007;14(5):662-73.
13. El Fadly A, Rance B, Lucas N, Mead C, Chatellier G, Lastic P-Y, et al. Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform. J Biomed Inform. 2011;44 Suppl 1:S94-102.
14. Nordo A, Eisenstein EL, Garza M, Hammond WE, Zozus MN. Evaluative Outcomes in Direct Extraction and Use of EHR Data in Clinical Trials. Stud Health Technol Inform. 2019;257:333-40.
15. Garza M, Myneni S, Nordo A, Eisenstein EL, Hammond WE, Walden A, et al. eSource for Standardized Health Information Exchange in Clinical Research: A Systematic Review. Stud Health Technol Inform. 2019;257:115-24.
16. Claerhout B, Kalra D, Mueller C, Singh G, Ammour N, Meloni L etal. Federated electronic health records research technology to support clinical trial protocol optimization: Evidence from EHR4CR and the InSite platform. J Biomed Inform. 2019;90:103090.

17. Topaloglu U, Palchuk MB. Using a Federated Network of Real-World Data to Optimize Clinical Trials Operations. JCO Clin Cancer Inform. 2018 Dec;2:1-10.
18. Rubinstein SM, Warner JL. CancerLinQ: Origins, Implementation, and Future Directions. JCO Clin Cancer Inform. 2018 Dec;2:1-7.
19. Fiteni F, Villanueva C, Bazan F, Perrin S, Chaigneau L, Dobi E et al. Long-term follow-up of patients with metastatic breast cancer treated by trastuzumab: impact of institutions. Breast. 2014 Apr;23(2):165-9.
20. Daniel C, Salamanca E. Les bases de données des hôpitaux. L'Entrepôt de données de l'AP-HP. In : Santé et intelligence artificielle. Cedric Villani, Bernard Nordlinger, éd. Paris : CNRS ; 1996.
21. Meystre SM, Friedlin FJ, South BR, Shen S, Samore MH. Automatic de-identification of textual documents in the electronic health record: a review of recent research. BMC Med Res Methodol, 2010 10(70).
22. Daniel C, Serre P, Orlova N, Bréant S, Paris N, Griffon N. Initializing a hospital-wide data quality program. The AP-HP experience. Comput Methods Programs Biomed. 2019;181:104804.

Tableau 1. Modalités d'accès aux données des EDS. L'exemple de l'AP-HP.

• **Vous êtes professionnel de santé de l'AP-HP et souhaitez réaliser une recherche « interne »** : une recherche est qualifiée d'« interne » si elle est menée à partir de données recueillies dans le cadre du suivi individuel des patients par les personnels assurant ce suivi et pour leur usage exclusif. o ***Exemple** : vous êtes médecin interniste exerçant à l'hôpital ou un professionnel de santé agissant sous sa responsabilité et souhaitez, à votre seul bénéfice, faire une étude (ex : indicateurs, taux de rémission, survie etc.) sur les patients que vous avez suivis au sein de votre service d'exercice.*
o Avec l'accord de votre chef de service, vous adressez au coordonnateur EDS de votre groupement hospitalier une demande d'accès aux données des patients pris en charge par le service. Pour ce type de recherches, aucun avis du CSE, ni aucune formalité CNIL ne sont requis. Le délai de mise à disposition des données est d'une à deux semaines. Pour les personnels assurant le suivi des patients l'accès aux données nominatives est possible. En tant que responsable de traitement vous devez inscrire le traitement de données concerné dans le registre des activités de traitement du groupement hospitalier.
• **Vous êtes chercheur ou acteur de la santé digitale, appartenant ou non à l'AP-HP, et souhaitez réaliser une recherche à plus grande échelle à partir des données de l'EDS AP-HP**
o Le kit de demande d'accès à l'EDS accessible sur le site de l'EDS de l'AP-HP* comporte le formulaire de demande d'accès que vous devez adresser au secrétariat du CSE accompagné du protocole de la recherche, de votre déclaration publique d'intérêts et de la liste des personnes destinataires des données dans le cadre de la recherche. Un data scientist au sein de l'unité de recherche clinique de votre groupement hospitalier peut vous aider dans cette démarche en réalisant notamment une étude de faisabilité de votre recherche. Après avis favorable du CSE, si les données nécessaires à la recherche sont disponibles au sein de l'EDS, le délai de mise à disposition de l'échantillon de données (sans compter le délai de contractualisation en cas de partenariat) est de 4 à 6 semaines.
• **Vous êtes chercheur ou acteur de la santé digitale, appartenant ou non à l'AP-HP, et souhaitez réaliser une recherche à partir des données de la base de données EDS-COVID**
o Une procédure accélérée a été mise en place dont les modalités sont accessible sur le site de l'EDS de l'AP-HP#. Après avis favorable du CSE, le délai d'accès à la base EDS-COVID est de quelques jours.
• **Dans tous les cas, tous les destinataires des données sont formés aux solutions de gestion et d'analyse des données et aux règles de protection et d'utilisation des données**

* https://eds.aphp.fr/demande-acces-donnees

https://eds.aphp.fr/sites/default/files/2020-03/COVID-19_EDS_APHP_V1.pdf