



**HAL**  
open science

## Quel impact du nouveau règlement européen sur la protection des données sur la recherche clinique et recommandations

Jacques Demotes-Mainard, Catherine Cornu, Aurelie Guerin, Pierre-Henri Bertoye, Romain Boidin, Serge Bureau, Jean-Marie Chretien, Cecile Delval, Dominique Deplanque, Claude Dubray, et al.

### ► To cite this version:

Jacques Demotes-Mainard, Catherine Cornu, Aurelie Guerin, Pierre-Henri Bertoye, Romain Boidin, et al.. Quel impact du nouveau règlement européen sur la protection des données sur la recherche clinique et recommandations. *Thérapie*, 2019, 74 (1), pp.17-29. 10.1016/j.therap.2018.11.008 . hal-02082395

**HAL Id: hal-02082395**

**<https://univ-rennes.hal.science/hal-02082395>**

Submitted on 16 Apr 2019

**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.

**ATELIERS DE GIENS 2018/RECHERCHE CLINIQUE****Quel impact du nouveau règlement européen sur la protection des données sur la recherche clinique et recommandations \***

Impact du RGPD en recherche clinique

Jacques Demotes-Mainard <sup>a</sup>, Catherine Cornu <sup>b,\*</sup>, Aurélie Guérin <sup>c</sup>, les participants à la table ronde « Recherche clinique » des Ateliers de Giens XXXIV, Pierre-Henri Bertoye <sup>d</sup>, Romain Boidin <sup>e</sup>, Serge Bureau <sup>f</sup>, Jean-Marie Chrétien <sup>g</sup>, Cécile Delval <sup>h</sup>, Dominique Deplanque <sup>i</sup>, Claude Dubray <sup>j</sup>, Luc Duchossoy <sup>k</sup>, Valérie Edel <sup>l</sup>, Cécile Fouret <sup>m</sup>, Ariane Galaup <sup>n</sup>, Frédérique Lesaulnier <sup>o</sup>, Mihaela Matei <sup>a</sup>, Florian Naudet <sup>p</sup>, Valérie Plattner <sup>q</sup>, Michel Rubio <sup>r</sup>, Fabrice Ruiz <sup>s</sup>, Sophie Sénéchal <sup>t</sup>, Tabassome Simon <sup>u</sup>, Anne Vidal <sup>v</sup>, Anne Viola <sup>w</sup>, Mireille Violleau <sup>x</sup>

<sup>a</sup> ECRIN, 75013 Paris, France

<sup>b</sup> Service de santé publique, hospices civils de Lyon, France, INSERM CIC1407, 69677 Bron, France ; UMR5558 université de Lyon, 69372 Lyon, France

<sup>c</sup> Pfizer, 75014 Paris, France

<sup>d</sup> CNRIPH, ministère des solidarités et de la santé, 75013 Paris, France

<sup>e</sup> Unicancer, 75013 Paris, France

<sup>f</sup> AP-HP, 75475 Paris, France

<sup>g</sup> CHU Angers, DRCl, 4900 Angers, France

<sup>h</sup> AIRLIQUIDE, 78350, Les Loges-En-Josas, France

<sup>i</sup> Univ. Lille, Inserm, CHU Lille, CIC, 1403, centre d'investigation clinique, 59000 Lille, France

<sup>j</sup> CHU de Clermont-Ferrand, centre de pharmacologie clinique, 63003 Clermont Ferrand, France

<sup>k</sup> Sanofi, 94250 Gentilly, France

<sup>l</sup> INDS, 94220 Charenton le Pont, France

<sup>m</sup> Medtronic, 92100 Boulogne-Billancourt, France

<sup>n</sup> Les entreprises du médicament (Leem), 75017 Paris, France

<sup>o</sup> INSERM, 75654 Paris, France

<sup>p</sup> Univ Rennes, CHU Rennes, Inserm, CIC 1414, 35000 Rennes, France

<sup>q</sup> DRCI, hospices civils de Lyon, France, 69002 Lyon, France

<sup>r</sup> CHU de Clermont Ferrand, direction qualité, 63003 Clermont Ferrand, France

<sup>s</sup> Clinsearch, 92245 Malakoff, France

<sup>t</sup> Astrazeneca, 92400 Courbevoie, France

<sup>u</sup> Service de pharmacologie clinique, plateforme de recherche clinique de l'Est parisien (URC-CRC-CRB), APHP, hôpital Saint Antoine, 75571 Paris, France

<sup>v</sup> CNIL, service santé, Paris, 75017 Paris, France

<sup>w</sup> Lysarc, centre hospitalier de Lyon, 69495 Pierre-Bénite, France

<sup>x</sup> Roche SAS, 92650 Boulogne-Billancourt, France

Texte reçu le 9 novembre 2018 ; accepté le 13 novembre 2018

\*Auteur correspondant. Centre d'investigation clinique, hôpital Louis Pradel, 28 avenue du Doyen Lépine, 69500 Bron, France.

Adresse e-mail : catherine.cornu@chu-lyon.fr (C. Cornu)

\*Les articles, analyses et propositions issus des Ateliers de Giens sont ceux des auteurs et ne préjugent pas des propositions de leur organisation.

## Résumé

La recherche clinique, menée sur des sujets humains ou à partir de leurs données, se trouve confrontée à des exigences contradictoires, avec d'une part le principe de science ouverte (transparence et partage des données), les possibilités offertes par les « big data » et la réutilisation des données issues de la recherche ou du soin, et d'autre part l'encadrement réglementaire et législatif qui évolue, avec notamment le règlement général sur la protection des données ou RGPD.

Une table ronde était organisée à Giens en Octobre 2018, afin d'identifier les zones de tension, les besoins de clarification et de fluidification, et de faire des recommandations visant à promouvoir la recherche clinique tout en assurant un haut niveau de protection des patients.

Après une mise au point effectuée en commun sur ces évolutions, les participants à la table ronde ont pu proposer des recommandations, dont les principales sont 1) de clarifier : ce qui est considéré comme données anonymisées, et ce que représente « l'intérêt public » au sens du RGPD ; 2) de poursuivre le développement des méthodologies de référence par la CNIL ; 3) de promouvoir l'utilisation secondaire des données, en facilitant une information et un consentement élargi des patients, en clarifiant les circonstances d'application de leurs droits de retrait, d'opposition, de façon à limiter le risque de biais ; 4) faciliter l'accès aux entrepôts de données par la mise à disposition de supports technologiques et méthodologiques. Enfin développer les échanges au niveau européen entre autorités sur les sujets de recherche, promouvoir la participation des autorités françaises à l'élaboration des codes de conduite, et la mise en place d'une procédure d'harmonisation volontaire pour coordonner les avis des autorités de protection des données, tout en s'assurant que les principaux documents soient disponibles en langue anglaise.

## MOTS CLÉS

Protection des données ; Textes réglementaires ; Recherche clinique ; *Big data* ; CNIL

## Abréviations

CERES : Comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé

CHU : centre hospitalier universitaire

CNIL : Commission nationale de l'informatique et des libertés

CNRIPH : Commission nationale des recherches impliquant la personne humaine

CPP : comité de protection des personnes

CSP : code de la Santé publique

DPD : délégué à la protection des données (*data protection officer* [DPO])

ECG : électrocardiogramme

EM : État-membre (de l'Union européenne)

FAIR : *findable, accessible, interoperable, reusable*

IA : intelligence artificielle

HIPAA : *Health insurance portability and accountability act*

INDS : Institut national des données de santé

MR : méthodologie de référence

NIR : numéro d'identification au répertoire

OMS : Organisation mondiale de la santé

PMSI : programme de médicalisation des systèmes d'information

RGPD : règlement général sur la protection des données

RT : responsable de traitement

SNIRAM : système national d'information inter-régimes de l'Assurance maladie

SNDS : système national des données de santé

ST : sous-traitant

## Introduction

Comme de multiples secteurs de la vie sociale, la recherche clinique est en passe de connaître une révolution digitale dont l'origine est double – accès à des données massives, et possibilité de réutiliser des données digitales.

L'avènement de techniques « -omiques » à haut débit et d'imagerie permettent de générer, pour chaque patient, des masses considérables de données digitalisées, et leur exploitation par des algorithmes pour stratifier les patients selon des profils de biomarqueurs complexes – c'est l'ambition de la médecine « personnalisée ».

Simultanément, la digitalisation des données de recherche clinique et des données issues du soin ou des systèmes de santé permet leur réutilisation en recherche pour des usages secondaires, qu'il s'agisse d'essais cliniques ou d'études observationnelles. Ces usages secondaires incluent la ré-analyse, des analyses secondaires ou des méta-analyses. Il est aussi possible de réutiliser des cohortes ou des registres pour réaliser des essais interventionnels [1-3]. Les informations issues des entrepôts de données hospitaliers peuvent être utilisés pour sélectionner les sites dans un essai clinique, pour sélectionner les patients, optimiser le design de l'étude, ou pour recueillir les données d'un essai clinique ou d'une étude en 'vie réelle'. L'utilisation de données issues d'objets connectés pour un essai clinique ou pour une étude en vie réelle va elle aussi devenir de plus en plus courante.

Enfin la politique de science ouverte, en application du principe *findable, accessible, interoperable, reusable* (FAIR) et dans la perspective d'améliorer la reproductibilité des résultats de la recherche, plaident en faveur du partage et de la réutilisation des données de recherche clinique.

Ces évolutions soulèvent des questions multiples en regard de la méthodologie, de la protection des données, de leur dé-identification et du risque de ré-identification, de l'accès et du partage de données, de l'information et la ré-information des personnes concernées, ainsi que des modalités de recueil du consentement [4]. Or l'encadrement réglementaire et législatif de ces questions connaît lui aussi des évolutions substantielles, en particulier avec la mise en application du RGPD (Tableau 1).

Ces sujets ont été abordés lors d'une table ronde organisée à Giens en Octobre 2018, afin d'identifier les zones de tension, les besoins de clarification et de fluidification, et de faire des recommandations visant à promouvoir la recherche clinique tout en assurant un haut niveau de protection des patients. Le programme recouvrait la nouvelle donne en matière de réglementation applicable à la recherche clinique (et ses conséquences sur l'information du patient, le consentement élargi, le retrait de consentement), les dispositifs de simplification et d'encadrement mis en place par la Commission nationale de l'informatique et des libertés (CNIL) et ses méthodologies de référence (MR), l'impact pour les

établissements publics et privés en termes de gouvernance, d'outils de conformité, le rôle du délégué à la protection des données (DPD°[Tableau 2]), le contexte européen et international, le partage des données, des aspects techniques concernant la sécurité et confidentialité des données, la minimisation (anonymisation vs. pseudonymisation) et le risque de ré-identification, les conditions pour l'hébergement de données de santé, l'agrégation de données non téléchargeables, le partage et la réutilisation pour la recherche clinique des données issues de la recherche (données d'essais cliniques - repositoires, sécurité, conditions d'accès et d'analyse secondaire, plan de partage des données) ou issues du système de santé, et l'utilisation de données issues d'objets connectés.

## Contexte

### Le contexte réglementaire

Le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, abrogeant la directive 95/46/CE (RGPD) est entré en application le 25 mai 2018 [5]. Les objectifs du RGPD sont d'harmoniser au niveau européen les règles en matière de protection des données, de renforcer et moderniser les droits des personnes sur leurs données personnelles, de responsabiliser les acteurs, et de faciliter les flux de données. La figure 1 rappelle les grands principes du traitement des données.

Dans le même temps, l'environnement de la recherche clinique évolue sur le plan réglementaire, avec l'entrée prochaine en application du règlement 2014/536 du Parlement européen et du Conseil relatif aux essais cliniques de médicaments à usage humain qui abroge la directive 2001/20/CE [6]. Dans le domaine du dispositif médical, un nouveau règlement a été adopté par le Parlement et le Conseil en mai 2017 (règlement 2017/746) [7] afin d'harmoniser les conditions d'autorisation et de conduite des investigations cliniques portant sur les dispositifs médicaux, ainsi que leur accès au marché.

L'utilisation secondaire des données, en dehors du protocole de l'essai clinique, devrait se conformer au droit applicable en matière de protection des données, notamment au RGPD et au droit national. Toutefois le règlement 2014/536, par son article 28, prévoit explicitement la possibilité pour le promoteur de demander au participant à l'essai clinique, au moment où ce dernier donne son consentement pour participer à l'essai, d'accepter que ses données soient réutilisées « *en dehors du protocole de l'essai clinique exclusivement à des fins scientifiques* » [6]. Ce consentement élargi, qui peut être retiré à tout moment et qui sert de base pour l'utilisation secondaire des données, s'appliquerait donc pour les essais cliniques de médicament. Ce même article précise néanmoins que la recherche scientifique qui exploite les données en dehors du protocole de l'essai clinique a lieu

conformément au droit applicable en matière de protection des données, et l'articulation entre le règlement essais cliniques et le RGPD fait l'objet actuellement de discussions au niveau européen.

En France, le législateur a ouvert l'accès aux données de santé, telles que les données de l'assurance maladie ou les données issues de l'activité des établissements de santé, favorisant ainsi la réalisation de recherches et d'études dans les domaines de la santé et de la prise en charge médico-sociale.

### Protection des données

Des dispositions particulières relatives au traitement des données de santé coexistent en France avec le RGPD. Un chapitre, au sein de la loi informatique et libertés modifiée, est dédié aux recherches, études et évaluations dans le domaine de la santé (chapitre IX, section 2) [8]. Certaines conditions doivent être respectées, telles que la finalité d'intérêt public de l'étude, l'avis d'un comité préalable ou l'information individuelle des personnes concernées. Les procédures de demande d'autorisation sont décrites dans la figure 2.

Afin d'alléger les démarches administratives des promoteurs de recherche, des procédures simplifiées ont été mises en place depuis 2006 par le biais des méthodologies de référence (MR). Ces MR, qui font désormais partie des référentiels, exigent un haut niveau de confidentialité et de protection des personnes. Elles dispensent les responsables de traitement de déposer une demande d'autorisation auprès de la Commission nationale de l'informatique et des libertés (CNIL). La démonstration de la conformité à une méthodologie de référence remplace le régime d'autorisation.

Enfin, la loi informatique et libertés s'applique au responsable de traitement situé en France (même s'il ne traite pas des données de personnes en France) et, au responsable de traitement situé hors de France s'il traite des données de santé de personnes résidant en France (Tableau 3).

### **Minimisation des données : anonymisation versus pseudonymisation**

Le RGPD s'applique à toute donnée à caractère personnel se rapportant à une personne physique identifiée ou identifiable. L'article 2 de la Loi n° 78-17 du 6 janvier 1978, dite loi informatique et libertés, clarifie que « pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne ». Donnée personnelle et donnée anonyme s'opposent donc par définition. S'il existe un lien permettant la ré-identification de la personne, la donnée est à considérer comme étant à caractère personnel. La donnée ne sera anonyme que dans la mesure où elle ne permet aucune ré-identification de la personne, y compris par croisement avec une nouvelle base de données. Cette donnée anonyme n'est pas concernée par la réglementation relative au traitement des données personnelles.



Travailler sur une donnée anonyme permettrait aux chercheurs de s'affranchir des dispositions du RGPD, telles que l'information des personnes, le maintien d'un niveau de sécurité adéquat, et la formalisation de tout traitement de données.

La réglementation américaine en matière de protection des données personnelles (HIPAA)<sup>1</sup> propose de « dé-identifier » les données personnelles afin de permettre leur partage sans contrainte, par l'effacement d'une liste de données définies comme identifiantes par nature, telles que les noms, codes postaux, dates de naissance, etc.

Le cadre réglementaire européen considère plutôt un principe de minimisation de la donnée (article 5.1.c du RGPD) afin de ne traiter que les données strictement nécessaires, sans préjuger des moyens pratiques mis en œuvre à cet effet.

La possibilité d'identification des personnes s'interprète en Europe au regard des moyens raisonnablement susceptibles d'être utilisés en prenant en considération l'ensemble des facteurs objectifs tel que le coût et le temps nécessaires à l'identification ainsi que les technologies disponibles au moment du traitement mais également de leur évolution. La notion d'anonymat doit être raisonnablement robuste dans le temps.

Toutefois, les méthodes d'anonymisation sont complexes et leur robustesse dans le temps difficile à apprécier [9-14]<sup>2</sup>. Il convient donc de considérer toute donnée de recherche en santé comme une donnée personnelle non-anonyme, jusqu'à preuve du contraire.

---

<sup>1</sup> HIPAA : Items à retirer pour dé-identifier des données de santé : (A) noms, (B) Subdivisions géographiques plus petites qu'un État, (C) Tous les éléments de date (sauf année) pour les dates qui sont directement liées à un individu, y compris la date de naissance, (D) Numéros de téléphone, (L) Identifiants du véhicule, (E) Numéros de fax, (M) Identificateurs d'appareil et numéros de série, (F) adresses e-mail, (N) localisateurs de ressources universelles Web (URL), (G) Numéros de sécurité sociale, (O) adresses de protocole Internet (IP), (H) Numéros de dossier médical, (P) Identifiants biométriques, y compris empreintes digitales et vocales, (I) Nombre de bénéficiaires du régime de santé, (Q) Photographies complètes et toute image comparable, (J) Numéros de compte, (R) Tout autre numéro, caractéristique ou code d'identification unique, (K) Numéro de certificat / licence.

<sup>2</sup> Peut-on rendre des données anonymes et comment faciliter leur partage pour la recherche ?

Plusieurs expériences de diffusion de données anonymisées en OpenData aux Etats-Unis (GIC 2002, AOL 2006, Netflix 2010) [9,10] ont conduit à des ré-identifications de personnes. En 2016, en Australie, une diffusion de données du Système national de santé a été interrompue suite à la découverte de possibilité de ré-identification [11,12] Dans le domaine sensible de la santé, la diffusion de données en accès totalement libre semble donc être une promesse difficile à tenir compte-tenu des risques pour la vie privée des personnes. Dans son analyse des techniques d'anonymisation le groupe de l'article 29 [9] considère qu'aucune technique ne garantit une anonymisation à 100% sans dégrader de façon importante les données. Une véritable anonymisation est un processus complexe faisant appel à une combinaison de techniques de généralisation (k-anonymat, l-diversité) et de randomisation (ajout de bruit, confidentialité différentielle,...) [13]. Ces techniques complexes diminuent la quantité d'information dans les données

## Accès aux bases médico-administratives, et entrepôts de données hospitaliers

### Système national des données de santé (SNDS)

La loi Française de modernisation du système de santé adoptée en janvier 2016 a consacré la création d'une base de données médico-administrative d'une richesse précieuse : le SNDS [15]. Il s'agit d'une base de données pseudonymisées rassemblant les consommations de soins de ville et en établissement de santé de l'ensemble de la population française, du programme de médicalisation des systèmes d'information (PMSI), du registre des causes de décès, et prochainement enrichie par des données sur le handicap et de données issues des assurances complémentaires de santé. L'accès à ces données est dorénavant largement ouvert à tout acteur, la contrepartie étant d'une part un renforcement des conditions assurant la confidentialité et l'intégrité des données et la traçabilité des accès, et d'autre part un retour vers le citoyen via la transparence des travaux menés à partir de ces données et des résultats en découlant.

Concernant les procédures d'accès, des procédures simplifiées (MR) ont d'ores et déjà vu le jour en juillet dernier. Si un traitement n'est pas conforme aux exigences des MR, une autorisation doit être demandée auprès de la CNIL (Figure 2).

Mais le SNDS est avant tout une formidable base « pivot » qui tirera toute sa richesse de sa mise en relation avec d'autres sources de données cliniques, environnementales, sociologiques. L'assouplissement des conditions d'utilisation du numéro d'identification au répertoire (NIR [numéro de sécurité sociale]) devrait faciliter la multiplication de ces mises en relation qui devront être traitées dans des environnements très sécurisés. À cet égard, les attentes sont fortes concernant les travaux issus de la mission de préfiguration sur le « *health data hub* » [16].

### Évolution du dossier médical

---

et la capacité à les réutiliser pour la recherche. Un intermédiaire à ces techniques d'anonymisation forte nous semble être la minimisation renforcée qui consiste à retirer les principaux éléments identifiants et semi-identifiants (liste des items HIPAA aux USA), à remplacer toutes les dates par des délais et à généraliser certaines variables lorsque cela est possible. Une base préparée en utilisant cette technique devrait pouvoir être partagée avec la communauté scientifique au travers d'une plateforme tierce garante des modalités d'accès et de sécurisation des données en utilisant des processus tels que ceux décrits par Ohmann et al. [3, 14]. Afin de concilier impératifs de transparence et exigence de protection des données personnelles, en addition de la mise en place d'outils de partage, il conviendrait de mettre en place des codes de conduite permettant de faciliter le partage des données de santé à des fins de recherche scientifique, en accord avec les considérants 26, 33 et 50, et en application des articles 14.5, 40 et 89 du RGPD.

Conçu dans des laboratoires universitaires au début des années 1980, le dossier médical informatisé, initialement destiné à la gestion des données de soins et la codification des diagnostics et des actes, a changé d'usage au fil de son industrialisation depuis les années 1990. Il est aujourd'hui principalement un outil de suivi de l'activité médicale et de facturation. Il en résulte qu'il est bien souvent une copie numérique de ce qu'était le dossier papier sans valeur ajoutée pour la recherche. Les données sont difficilement exploitables du fait de la faible interopérabilité, de la multiplicité des applications qui le compose, et du peu de structuration d'une grande partie des données. L'avènement de technologies permettant de traiter des masses importantes de données et l'utilisation croissante de formats de données standardisés offrent aujourd'hui la possibilité de construire des entrepôts de données de santé à l'échelle d'un établissement, voire d'un territoire, pour agréger toutes les données d'un patient. Les nouvelles technologies de recherche et d'indexation *full-text* permettent, grâce à l'enrichissement sémantique *via* les ontologies médicales, d'indexer et de requêter des documents textuels non structurés tels que les observations médicales et les comptes rendus d'examen ou d'hospitalisation, qui constituent une source majeure d'information. Ces données cliniques hospitalières sont complémentaires des données du système national d'information interrégimes de l'Assurance maladie (SNIIRAM) qui apportent des informations précieuses sur la consommation de soins et le parcours des patients qui ont été suivis à l'hôpital.

À l'échelle d'un établissement ou d'un territoire, l'accès à ces données doit être encadré (éthique, réglementation, méthodologie, information des patients, transparence des traitements). La CNIL, qui délivre les autorisations nécessaires, attache une importance particulière à la sécurisation des données, à l'information des patients et à la gouvernance de ces entrepôts. La solution mise en place au sein du plus important centre hospitalier universitaire (CHU) de France (Paris) et le projet eHOP du CHU de Rennes comprennent la mise en place d'une structure « tiers de confiance » chargée de l'analyse des demandes, de leur traitement et de la mise à disposition des données des patients nécessaires à un projet, dans le respect de l'éthique et de la réglementation.

Demain, ces données viendront s'enrichir des données apportées par les patients (*self-quantified*), de données génétiques (*-omique*) et de données recueillies en routine (données des moniteurs, ventilateurs, électrocardiogramme [ECG]). Les technologies d'intelligence artificielle (IA) actuellement en plein développement pourront tirer pleinement parti de ces données, en particulier la fouille de données non-structurées présentes dans les entrepôts de données.

### Les données issues d'objets connectés

Les essais cliniques, comme les études en vie réelle, peuvent également être réalisés en utilisant des données issues d'objets connectés et d'outils de télé-suivi. Les pathologies permettant ce type de suivi sont de plus en plus nombreuses (diabète, insuffisance cardiaque, rythmologie, etc). Ces modes de recueil sont pris en

compte dans les MR, cependant de nombreux télé-suivis ont été initiés sans envisager la récupération des données à des fins de recherche clinique, voire la récupération de données à des fins d'amélioration de la performance algorithmique. Se pose donc de façon aiguë la question des modalités d'information des patients et de recueil de la non-opposition, ainsi que celle de la sécurité, de la confidentialité et de la qualité des données issues d'objets connectés utilisés dans le cadre de la recherche.

La réalisation d'études en vie réelle utilisant des bases de données de soins, de cohortes et essais cliniques

La possibilité de réaliser des études utilisant les bases de données médico-administratives transforme déjà le paysage de la recherche clinique. Ainsi la cohorte Constances, représentative de la population française, inclut 200 000 volontaires âgés de 18 à 69 ans, et combine : 1) des visites dans un centre d'examen de santé de l'une des caisses primaires de l'assurance maladie participantes, afin de bénéficier d'un examen médical, paraclinique, biologique et, pour ceux âgés de 45 ans et plus, de tests cognitifs et physiques ; 2) une biobanque ; 3) l'utilisation de données issues des bases médico-administratives (caisse nationale d'assurance vieillesse, caisse d'assurance maladie, hôpitaux et professionnels de santé) ; et 4) des questionnaires portant sur des thématiques variées en relation avec l'état de santé, régulièrement remplis par les volontaires (Tableau 4) [17,18].

De même le projet iVasc (cofinancé par des fonds publics et industriels), comporte une cohorte (FRENCHIE) de patients hospitalisés pour infarctus du myocarde, dont le suivi est assuré grâce aux bases de données médico-administratives, et permet également de nicher des essais interventionnels au sein de la cohorte, dont deux sont déjà planifiés et financés - sur la santé dentaire et les troubles du sommeil.

### **La nécessité scientifique et éthique du partage de données**

L'application du RGPD devra se faire dans le cadre plus global du mouvement de la science ouverte qui révolutionne les normes jusqu'alors admises par les scientifiques, en imposant 3 valeurs scientifiques clefs : la transparence, l'ouverture et la reproductibilité. Le partage des données a une justification éthique. En participant aux essais les personnes prennent des risques, pour des bénéfices incertains. Il en résulte un contrat implicite avec l'obligation que les données issues de ces essais entraînent le plus grand bénéfice pour la société dans son ensemble. Pour maximiser ces bénéfices, le Consortium international des éditeurs de journaux médicaux (ICMJE) prévoit qu'à compter de janvier 2019 : 1) les auteurs rapportant les résultats d'essais cliniques détaillent leurs intentions quant au partage des données dé-identifiées et des métadonnées associées [19] et 2) un plan de partage des données soit défini et enregistré avant le début de l'essai clinique, en même temps que l'essai sur un registre type ClinicalTrials.gov [20-22]. Plusieurs revues, dont le BMJ et les revues du groupe PloS,

ont devancé ce mouvement en incitant depuis plusieurs années les auteurs à fournir leurs données. Il n'en reste pas moins que ces exigences de transparence peuvent être rendues difficiles par la mise en application du RGPD.

Comment gérer la contradiction entre protection des données personnelles et transparence de la recherche / partage des données ?

Il existe en effet une apparente contradiction entre le besoin de renforcer la protection des données personnelles et celui, tout aussi éthique, de partager les données issues du soin ou de la recherche pour améliorer la connaissance et *in fine* les soins pour les patients. Certaines dispositions issues du RGPD concernant l'information et le consentement du patient vis-à-vis des utilisations secondaires pourraient fragiliser ces avancées. Ces dispositions pourraient faire courir un risque quant à la validité scientifique des données réutilisées à des fins de recherche. En particulier (sauf à utiliser des données anonymes qui échappent au RGPD) la possibilité de ne pas consentir à un usage secondaire, de ne consentir que pour des usages restreints (pour une institution ou une pathologie données), et plus encore la possibilité de retirer son consentement à un usage secondaire risquent d'augmenter le risque de biais si de nombreux patients faisaient valoir ce droit – en particulier si la satisfaction du patient lors de son hospitalisation, ou lors d'un essai, interfère avec sa décision de retrait du consentement pour une utilisation secondaire.

Une enquête auprès de 771 participants à des essais cliniques aux États-Unis a montré que 93 % des répondants sont susceptibles de permettre le partage de leurs propres données avec des scientifiques universitaires, et 82 % avec des scientifiques d'entreprises à but lucratif [23] - un taux d'acceptation meilleur que pour le partage des données hospitalières ou des échantillons biologiques.

### **Impact du RGPD sur la recherche clinique**

La référence au RGPD s'impose ainsi à tous les aspects de la recherche clinique, depuis l'information des patients, qui doit désormais comporter la référence au RGPD et les éléments d'information qu'il préconise, jusqu'à la publication des résultats, où il peut entrer en contradiction avec les exigences de transparence et de partage de données. C'est pour atténuer ces tensions que nous proposons les recommandations suivantes.

## Recommandations

Ces recommandations s'inscrivent dans trois directions : l'exercice du droit des personnes, la fluidification des procédures administratives, et l'ouverture internationale.

### Recommandations concernant l'exercice des droits des personnes au regard leurs données

#### RGPD et information des patients

L'entrée en vigueur du RGPD renforce l'obligation d'information des patients, qui sont par ailleurs, comme tout citoyen, saturés de messages provenant de fournisseurs de services, commerces, etc. sur l'actualisation du RGPD, ce qui peut brouiller le message sur la participation à la recherche. Pour éviter cet écueil, la table ronde recommande la mise à disposition d'une **version commune de formulaire d'information des patients** quant au contenu (modèle de document proposé par la Commission nationale des recherches impliquant la personne humaine [CNRIPIH]). Une proposition de ligne directrice précisant quelle information donner à quels patients (nouveaux, patients déjà inclus, patients ayant terminé une étude), à quel moment, et par quel acteur a été diffusée le 25 mai 2018 par la CNRIPIH, et est en cours de révision.

#### Exercice des droits des patients

Le droit à l'effacement et le droit à l'opposition peuvent entrer en divergence avec l'exigence méthodologique de pratiquer les analyses sur des données exhaustives, lors de l'utilisation princeps ou la réutilisation secondaire des données. En effet, les données manquantes constituent une menace pour la validité des estimations de l'effet du traitement dans les études contrôlées randomisées. Afin d'éviter les biais, il est recommandé de faire une analyse en intention de traiter, incluant tous les patients randomisés dans le groupe dans lequel ils ont été randomisés. Les taux de données manquantes doivent être précisés, et des méthodes spécifiques de prise en compte des données manquantes doivent être utilisées et explicitées, mais ces méthodes ne permettent pas d'éliminer totalement les biais liés aux données manquantes [24,25]. Cette exigence méthodologique figure également dans le règlement 2014/536 du Parlement européen et du Conseil relatif aux essais cliniques de médicaments à usage humain qui précise (Art. 28.3) « que le retrait du consentement éclairé n'a pas d'incidence sur les résultats des activités déjà menées, telles que la conservation et l'utilisation des données obtenues sur la base du consentement éclairé avant que celui-ci ne soit retiré ». Nous recommandons de **clarifier dans quelles circonstances ce droit de retrait s'applique à l'utilisation secondaire** (données de sécurité et d'efficacité), comment informer les patients des conséquences dommageables pour les conclusions de l'étude, dans quelles conditions permettre à un patient de refuser de façon différenciée et en connaissance de cause à la poursuite du traitement à l'étude, du suivi prévu par le protocole, du recueil de données lors du suivi habituel pour les

besoins de l'étude, voire à l'effacement de ses données. Une proposition des compléments à apporter à l'information-patient devrait être faite par la CNIL/CNRIPH.

Les conditions du droit d'opposition posent également question : les patients ont un droit d'opposition au traitement de leurs données mais les aspects pratiques de l'exercice de ce droit ne sont pas simples. **Il faudrait préciser et clarifier auprès de qui ces droits s'exercent** : en première intention l'investigateur (seule personne connue du patient et qui connaît son identité) et/ou le DPD du promoteur qui peut être destinataire de l'identité de la personne qui souhaite le contacter ; le DPD est soumis à une obligation de confidentialité ou de secret professionnel en ce qui concerne l'exercice de ses missions.

Le droit à la portabilité figure dans le RGPD, lorsque le fondement juridique est le consentement ou le contrat, pour permettre à des patients de disposer de leurs données médicales afin d'établir une continuité dans leur suivi médical lorsqu'ils changent de lieu de soins. Cette disposition est peu pertinente pour les données recueillies spécifiquement pour la recherche. L'explication donnée aux patients sur ce droit devient inintelligible à moins de détailler, pour une recherche donnée, comment ce droit pourrait s'appliquer. **Une dérogation à ce droit devrait être faite pour les données de la recherche** qui ne sont pas des données utiles au soin.

Réutilisation des données

- Les MR définissent la possibilité d'usage secondaire, avec information au patient. Cette question de l'information peut se poser de façon prospective, lorsque l'on planifie une recherche, pour anticiper les utilisations ultérieures, ou lorsqu'une base de données précédemment constituée est réutilisée. La recommandation est de faire en sorte que **l'information initiale du patient englobe les recherches ultérieures**. Pour qu'il soit le moins possible limitant, tout en préservant la confidentialité des données, **l'information initiale sur les réutilisations devrait porter sur les modalités de l'accès aux données plutôt que sur la thématique des recherches ultérieures** : qui peut accéder aux données ? Quelles données (quelle anonymisation) ? Selon quelles modalités : accès contrôlé (via une requête analysée par un comité scientifique), possibilité d'agréger des données non-téléchargeables, partage après anonymisation ? Quel repository sera utilisé, quelle est sa gouvernance ?  
Ainsi la **MR-004 pourrait être mise à jour avec un autre aménagement de l'information individuelle** : « Lorsque la recherche correspond à des critères clairs dont la personne concernée a pu prendre connaissance en amont concernant l'utilisation des données la concernant en fonction des finalités, des modalités d'accès aux données et des acteurs concernés et des mesures de protection des données appliquées. »
- Promouvoir la participation des **représentants de patients dans la gouvernance des repositories** permettant l'utilisation secondaire des données.

- **Valoriser le partage des données de recherche** : il serait important que les indicateurs de valorisation de la recherche incluent ce partage, les publications secondaires issues de ce partage, le nombre des accès à une base de données etc. Le RGPD maintient l'exigence d'une information individuelle des patients quant à l'utilisation secondaire de leurs données personnelles, mais il existe des possibilités d'exception, notamment *si la fourniture de telles informations se révèle impossible ou exigerait des efforts disproportionnés, compromettrait gravement la réalisation des objectifs dudit traitement*. Dans la mesure où l'absence d'information des personnes rend la recherche non éligible aux méthodologies de référence (MR), l'analyse de ces situations est faite au cas par cas par la CNIL, sur la base du dossier fourni par le demandeur, dans le cadre d'une demande d'autorisation (Chapitre IX de la loi informatique et libertés). La recommandation de la table ronde est que la CNIL mette à disposition: 1) une ligne directrice précisant les **critères de l'octroi de ces exceptions**, et 2) des **exemples de ce qui peut être considéré comme un « effort disproportionné »** (guidance européenne « *Transparency* » [22]).

### Recommandations concernant la fluidité des procédures administratives

- Il importe de **poursuivre le développement des méthodologies de référence**, et **d'éviter de les limiter à des acteurs identifiés** (telle la MR 005 qui s'applique aux établissements de santé, et exclut *de facto* les universités et organismes de publics de recherche).
- Il importe de proposer un cadre de **définition de la finalité d'intérêt public**, condition nécessaire pour le traitement des données à des fins de recherche en France.
- Des fiches techniques doivent être éditées pour servir de **support aux chercheurs qui veulent accéder aux différents entrepôts de données**, notamment du SNDS, incluant un inventaire des bases existantes, et leurs modalités d'accès.

En particulier, il convient de **clarifier ce qui est considéré comme données anonymisées ou pseudonymisées**, proposer des solutions validées d'anonymisation, des outils informatiques et des formations pour les data managers permettant d'anonymiser des données.

- Afin de favoriser les appariements et rendre possible le chaînage, supprimer la contradiction qui veut que **l'utilisation du NIR soit désormais possible** sans décret en Conseil d'État, sauf **s'il est issu d'un dossier médical**.

### Recommandation liée à l'annexe unique RGPD pour la convention unique : nécessité d'un contrat entre responsable de traitement et sous-traitant

Le RGPD a redéfini les responsabilités et obligations de l'ensemble des acteurs intervenant dans le traitement de données à caractère personnel. Le responsable de traitement (« RT ») est celui qui détermine,



seul ou conjointement avec d'autres, la finalité et les moyens du traitement mis en œuvre. Le sous-traitant (« ST ») est celui qui traite des données à caractère personnel pour le compte du RT. Il a de nouvelles obligations au regard du RGPD, sa responsabilité propre étant susceptible d'être engagée en cas de manquement.

La convention unique a été introduite par la Loi de modernisation de notre système de santé (« LMSS ») du 26 janvier 2016, dans le but d'améliorer les délais de mise en place des essais cliniques dans les hôpitaux Français, élément important dans l'attractivité de la France de la recherche clinique internationale [26]. Le modèle de convention unique, devant être utilisé, a été fixé par l'arrêté du 16 novembre 2016, prévu à l'article R. 1121-4 du code de la Santé publique (CSP), antérieurement au RGPD. Son champ d'application est la recherche impliquant la personne humaine, de type 1° et 2°, à finalité commerciale dans les établissements, maisons et centre de santé.

Dans le cadre de la convention unique, le RT au titre du RGPD est le promoteur d'une recherche à finalité commerciale impliquant la personne humaine. Le ST est l'établissement, maison ou centre de santé signataire de la convention unique.

L'article 28 du RGPD définit le cadre contractuel régissant les relations entre le RT et le ST. Il ressort de cet article que le RT doit s'assurer par contrat que le ST remplit toutes les conditions et offre toutes les garanties appropriées en termes de protection des données à caractère personnel. L'article 28.3 du RGPD précise notamment les clauses à faire figurer dans le contrat – il s'agit d'une liste non exhaustive.

L'article 28.3 du RGPD dispose ainsi que « Le traitement par un sous-traitant est régi par un contrat ou un autre acte juridique au titre du droit de l'Union ou du droit d'un État membre, qui lie le sous-traitant à l'égard du responsable du traitement, définit l'objet et la durée du traitement, la nature et la finalité du traitement, le type de données à caractère personnel et les catégories de personnes concernées, et les obligations et les droits du responsable du traitement ».

Par ailleurs, la méthodologie de référence MR-001 adoptée par la CNIL, qui encadre les traitements comprenant des données de santé et présentant un caractère d'intérêt public, réalisés dans le cadre de recherches nécessitant le recueil du consentement, reprend l'obligation d'établir un contrat entre le RT et le ST.

**Il est donc nécessaire d'introduire une clause de sous-traitance au sein de la convention unique afin d'assurer sa mise en conformité avec le RGPD. Une mise à jour sur ce point pourrait être envisagée via une modification du modèle de convention unique par arrêté du ministère de la Santé et ce, indépendamment des discussions en cours sur plusieurs autres problématiques relatives à la mise en œuvre de la convention unique.**

## Améliorer les supports techniques

- Dans le cadre des MR un paragraphe sur les modalités de **transmission des données individuelles aux peer-reviewers** des journaux scientifiques pour mettre à disposition des sites sécurisés correspondant à ce cahier des charges pour les chercheurs est prévu.
- Le répertoire INDS, où sont enregistrées les études conduites à l'aide du SNDS, devrait satisfaire les **spécifications pour les registres de l'Organisation mondiale de la santé (OMS)**[27], obligatoire réglementairement et pour publier. Ceci éviterait aux auteurs d'avoir à enregistrer une deuxième fois leurs études dans un registre conforme, accepté par les éditeurs de revues biomédicales. Ces registres doivent répondre à plusieurs critères et inclure le jeu minimal de 20 items [28].
- Certaines études nécessitent l'enregistrement des coordonnées des patients d'études cliniques, à des fins de remboursements de frais par exemple. Or la CNIL recommande une étanchéité entre les bases de données contenant les données personnelles de santé et les bases contenant les coordonnées personnelles pour le remboursement. Cette étanchéité concerne aussi les CHU promoteurs. Les CHU pourraient déclarer comme sous-traitants des services administratifs qui possèdent les coordonnées des patients pour procéder à leur remboursement. **Il faudrait clarifier si cette organisation peut rentrer dans le cadre des MR.**
- L'utilisation des données du SNDS implique un cadre réglementaire et technique précis. Ces exigences s'appliquent aussi aux études ou traitements informatiques dérivés de ces recherches, appelés systèmes-fils, ce qui complique la réutilisation en cascade de données enrichies par le SNDS. La législation peut en l'état donner lieu à une interprétation restrictive, selon laquelle le SNDS peut être utilisé pour des études, recherches ou évaluations, ce qui interdirait la constitution d'entrepôts de données appariés avec le SNDS. En outre, ces contraintes ne sont plus pertinentes lorsque les données du SNIIRAM utilisées dans les systèmes fils deviennent minimes. Il serait utile de donner une **définition précise de ce qu'est un système fils**.

## Favoriser la coopération internationale

- L'harmonisation à travers l'Europe des règles et procédures concernant la protection des données dans le cadre des recherches dans le domaine de la santé devrait être mieux documentée. A titre d'exemple, il est difficile à ce jour d'avoir une visibilité sur l'état d'intégration du RGPD dans chacun des pays européens et des obligations qui incombent au promoteur d'un essai clinique multinational. Par conséquent, une évaluation devrait être faite, une **analyse comparée européenne des législations nationales** et des exigences associées de protection des données de recherche en santé devrait être menée. Une mise au programme H2020 de cette proposition pourrait être envisagée.

- En cas de différences significatives entre les États-membres, une **procédure d'harmonisation volontaire pourrait être proposée. Celle-ci aurait pour objet de coordonner les avis des autorités de protection des données** des différents États-membres. Concernant les obligations vis-à-vis des CPP, afin de faciliter les projets internationaux il serait nécessaire d'**informer les promoteurs étrangers** sur la façon de procéder et rendre les documents disponibles. La CNRIPH s'engage à travailler sur ce sujet.
- Au regard du RGPD, les associations et les autres entités représentatives peuvent préparer des codes de conduite dans le but de préciser de quelle manière le RGPD s'applique à certains secteurs, notamment celui de la recherche scientifique. Ainsi faudra-t-il **renforcer l'implication des acteurs et autorités français dans l'élaboration de codes de conduite** concernant le traitement de données dans le cadre des essais cliniques, en particulier pour les traitements multinationaux. Un tel code est en cours d'élaboration, impliquant la collaboration de divers partenaires européens [29].
- La CNIL a mis en place des MR, qui ont beaucoup simplifié les démarches concernant la protection des données d'études cliniques. Ces outils et procédures ont souvent été traduits en anglais pour les usages internes de différents industriels. Une **traduction « officielle » des MR** devrait être fournie par la CNIL, et plus de visibilité devrait être faite sur leur existence pour informer les promoteurs étrangers.
- Enfin la **certification « hébergeur de données de santé »** en France requiert une procédure spécifiquement française, sur la base d'un référentiel fondé sur le respect de normes internationales (ISO 27001, 20000, 27018, 27017). L'évolution vers un système plus interopérable est recommandée, afin de ne pas isoler la France quant à l'usage des données de santé.

## Conclusions

La recherche clinique, menée sur des sujets humains ou à partir de leurs données, apparaît écartelée entre d'une part l'exigence de transparence et de partage des données caractéristiques de la science ouverte, les possibilités offerts par les « *big data* » (-omics), ou par la possibilité de réutiliser les données (données hospitalières, données des bases de données de santé, données issues de la recherche – cohortes et registres), et d'autre part les contraintes, encore imparfaitement définies, dérivées du RGPD.

Il s'agit pour tous les acteurs impliqués d'un sujet brûlant. La France doit éviter de s'isoler dans un système réglementaire faiblement interopérable et peut, au prix de quelques adaptations, réussir à protéger les données de ses citoyens tout en fluidifiant l'activité de recherche clinique. Les principales recommandations qui résultent de cette table ronde sont en premier lieu de clarifier ce qui est considéré comme données anonymisées (échappant ou RGPD) ou pseudonymisées (données personnelles, couvertes par le RGPD). Une clarification de ce qui est considéré comme d'intérêt public serait également bienvenue,

dans la mesure où il s'agit d'une condition de l'utilisation des données à des fins de recherche en France. Il convient aussi de poursuivre le développement par la CNIL des méthodologies de référence, et d'éviter de les limiter à des acteurs identifiés.

Afin de promouvoir l'utilisation secondaire des données, il importe que le consentement initial du patient englobe les recherches ultérieures, et décrive les modalités de l'accès aux données plutôt que de le restreindre à des recherches ultérieures sur une thématique limitée. Il convient de clarifier dans quelles circonstances le droit de retrait s'applique à l'utilisation secondaire, car il peut introduire des biais dans les analyses secondaires. Les chercheurs qui veulent accéder aux différents dépôts et entrepôts pour réutiliser des données doivent pouvoir bénéficier d'un support technologique et méthodologique par une infrastructure dédiée.

Enfin il faut éviter d'isoler la France dans le concert européen et international, en pratiquant un benchmark européen sur les législations et réglementations touchant la protection des données personnelles en recherche clinique, en développant les échanges entre autorités au niveau européen sur les sujets de recherche, en promouvant la participation des autorités françaises à l'accompagnement des codes de conduite, en proposant une procédure d'harmonisation volontaire pour coordonner les avis des autorités de protection des données, et en s'assurant que les principaux documents soient disponibles en langue anglaise.

#### **Déclaration de liens d'intérêts**

Les auteurs n'ont pas de liens d'intérêts à déclarer.

## Références

- [1] Ioannidis JP, Adami HO. Nested randomized trials in large cohorts and biobanks: studying the health effects of lifestyle factors. *Epidemiology* 2008 Jan;19(1):75-82.
- [2] Lauer MS, D'Agostino RB Sr. The randomized registry trial--the next disruptive technology in clinical research? *N Engl J Med* 2013 Oct 24;369(17):1579-81. doi: 10.1056/NEJMp1310102.
- [3] Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and Reuse of Individual Participant Data from Clinical Trials: Principles and Recommendations . *BMJ Open* 2017 Dec 14;7(12):e018647. doi: 10.1136/bmjopen-2017-018647.
- [4] . Chatellier G, Varlet V, Blachier-Poisson C, participants of Giens XXXI, Round Table No. 6. Big data and open data: what kind of access should researchers enjoy. *Thérapie*. 2016 Feb;71(1):97-105, 107-14.
- [5] CNIL. Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016. Mai 2018. <https://www.cnil.fr/fr/reglement-europeen-protection-donnees> [Consulté le 23 novembre 2018]
- [6] Journal officiel de l'Union européenne. EUR-Lex. Règlement 2014/536 du Parlement européen et du Conseil relatif aux essais cliniques de médicaments à usage humain et abrogeant la directive 2001/20/CE. Avril 2014. <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ:L:2014:158:TOC> [Consulté le 23 novembre 2018].
- [7] Journal officiel de l'Union européenne. Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (Texte présentant de l'intérêt pour l'EEE. ). Avril 2017. <http://data.europa.eu/eli/reg/2017/745/oj/fra> [Consulté le 23 novembre 2018]
- [8] CNIL. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Mai 2018. <https://www.cnil.fr/fr/loi-78-17-du-6-janvier-1978-modifiee> [Consulté le 23 novembre 2018]
- [9] Commission européenne. Article 29 Newsroom - News overview - European Commission. Mai 2018. <http://ec.europa.eu/newsroom/article29/news-overview.cfm> [Consulté le 23 novembre 2018]
- [10] Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*, Vol. 57, p. 1701, 2010; U of Colorado Law Legal Studies Research Paper No. 9-12. <https://papers.ssrn.com/abstract=1450006> [Consulté le 23 novembre 2018]
- [11] Culnane C, Rubinstein B, Teague V. Health data in an open world. Cornell university library. Decembre 2017. arXiv:1712.05627 [cs.CY] <https://arxiv.org/abs/1712.05627> [Consulté le 23 novembre 2018]
- [12] Australian government. Office of the Australian information commissioner. Publication of MBS/PBS Data - Office of the Australian information commissioner (OAIC). Mars 2018.

<https://www.oaic.gov.au/privacy-law/commissioner-initiated-investigation-reports/publication-of-mbs-pbs-data>. [Consulté le 23 novembre 2018]

[13] Sweeney L. K-anonymity: a model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 2002;10 (5):557-70.

[https://epic.org/privacy/reidentification/Sweeney\\_Article.pdf](https://epic.org/privacy/reidentification/Sweeney_Article.pdf) [Consulté le 23 novembre 2018 (14 pp.)]

[14] Ohmann C, Canham S, Banzi R, Kuchinke W, Battaglia S. Classification of Processes Involved in Sharing Individual Participant Data from Clinical Trials. Version 2. F1000Res. 2018 Feb 1 [revised 2018 Apr 20];7:138.

doi: 10.12688/f1000research.13789.2. eCollection 2018.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5861517/> [Consulté le 23 novembre 2018]

[15] Système national des données de santé. Septembre 2018. <https://www.snds.gouv.fr/SNDS/Accueil> (Consulté le 23 novembre 2018)

[16] Cuggia M, Polton D, Wainrib B, Combes S. Rapport *health data hub*. Octobre 2018. <https://www.ladocumentationfrancaise.fr/rapports-publics/184000690/index.shtml> [Consulté le 23 novembre 2018]

[17] Zins M, Goldberg M, et CONSTANCES team. « The French CONSTANCES Population-Based Cohort: Design, Inclusion and Follow-Up ». *Eur J Epidemiol* 2015 Dec;30(12):1317-28. <https://doi.org/10.1007/s10654-015-0096-4>. [Consulté le 23 novembre 2018]

[18] Feral-Pierssens AL, Carette C, Rives-Lange C, Matta J, Goldberg M, Juvin P, et al. Obesity and emergency care in the French CONSTANCES cohort. *PLoS One* 2018 Mar 26;13(3):e0194831. doi: 10.1371/journal.pone.0194831.

[19] Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, et al. Sharing clinical trial data: a proposal from the International committee of medical journal editors. *Lancet* 2016;387(10016) :PE9-E11. [https://doi.org/10.1016/S0140-6736\(15\)01279-9](https://doi.org/10.1016/S0140-6736(15)01279-9).

[20] Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the International committee of medical journal editors. *PLoS Medicine* 2017;14(6):e1002315. <https://doi.org/10.1371/journal.pmed.1002315>.

[21] Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing Raw Clinical Data for Publication: Guidance for Journal Editors, Authors, and Peer Reviewers. *Trials* 2010 Jan 29;11:9. doi: 10.1186/1745-6215-11-9.

[22] Bonini S, Eichler HG, Wathion N, Rasi G. Transparency and the European medicines agency--sharing of clinical trial data. *N Engl J Med*. 2014 Dec 25;371(26):2452-5. doi: 10.1056/NEJMp1409464.

[23] Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. *N Engl J Med* 2018 Jun 7;378(23):2202-11. doi: 10.1056/NEJMs1713258.

- [24] European medicines agency. Guideline on missing data in confirmatory clinical trials. Juillet 2010. [https://www.ema.europa.eu/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials\\_en.pdf](https://www.ema.europa.eu/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf) [Consulté le 23 novembre 2018 (12 pp.)]
- [25] Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med* 2013 Sep 20;86(3):343-58. eCollection 2013 Sep.
- [26] Galaup A, Barthélémy P, Pouletty-Lefebvre B, Béhier JM, Zetaloui J, Borel T. Attractivité de la France pour la recherche clinique internationale : résultats de la 8<sup>e</sup> enquête du Leem. *Thérapie* 2018;73(5):367-76.
- [27] World health organization. International clinical trials registry platform. Primary registries. 2018. <http://www.who.int/ictrp/network/primary/en/index.html> [Consulté le 23 novembre 2018]
- [28] World health organization. International clinical trials registry platform. WHO data set. 2018. <http://www.who.int/ictrp/network/trds/en/index.html> [Consulté le 23 novembre 2018]
- [29] BBMRI-ERIC. GPRD code of conduct. 2018. <http://www.bbmri-eric.eu/BBMRI-ERIC/gdpr-code-of-conduct> [Consulté le 23 novembre 2018]

References

1

Accepted Manuscript