



HAL
open science

How the new European data protection regulation affects clinical research and recommendations?

Jacques Demotes-Mainard, Catherine Cornu, Aurelie Guerin, Pierre-Henri Bertoye, Romain Boidin, Serge Bureau, Jean-Marie Chrétien, Cécile 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.. How the new European data protection regulation affects clinical research and recommendations?. *Thérapie*, 2019, 74 (1), pp.31-42. 10.1016/j.therap.2018.12.004 . hal-02082394

HAL Id: hal-02082394

<https://hal-univ-rennes1.archives-ouvertes.fr/hal-02082394>

Submitted on 22 Oct 2021

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

THERAPIE

HEADING: GIENS WORKSHOPS 2018/CLINICAL RESEARCH

Epub ahead of print puis numéro 1 (janv fév) 2019

How the new European data protection regulation affects clinical research and recommendations*

Impact of the GDPR on clinical research

Jacques Demotes-Mainard ^a, Catherine Cornu ^{b,c,*}, Aurélie Guérin ^d, the participants of Giens XXIV Round Table « Clinical research », Pierre-Henri Bertoye ^e, Romain Boidin ^f, Serge Bureau ^g, Jean-Marie Chrétien ^h, Cécile Delval ⁱ, Dominique Deplanque ^j, Claude Dubray ^k, Luc Duchossoy ^l, Valérie Edel ^m, Cécile Fouret ⁿ, Ariane Galaup ^o, Frédérique Lesaulnier ^p, Mihaela Matei ^a, Florian Naudet ^q, Valérie Plattner ^r, Michel Rubio ^s, Fabrice Ruiz ^t, Sophie Sénéchal-Cohen ^u, Tabassome Simon ^v, Anne Vidal ^w, Anne Viola ^x, Mireille Violleau ^y

^a ECRIN, 75013 Paris, France

^b Service de santé publique, hospices civils de Lyon, France, INSERM CIC1407, 69677 Bron, France

^c UMR5558 université de Lyon, 69372 Lyon, France

^d Pfizer, 75014 Paris, France

^e CNRIPH, ministère des solidarités et de la santé, 75013 Paris, France

^f Unicancer, 75013 Paris, France

^g AP-HP, 75475 Paris, France

^h CHU Angers, DRCL, 4900 Angers, France

ⁱ AIRLIQUIDE, 78350, Les Loges-En-Josas, France

^j Univ. Lille, Inserm, CHU Lille, CIC, 1403, centre d'investigation clinique, 59000 Lille, France

^k CHU de Clermont-Ferrand, centre de pharmacologie clinique, 63003 Clermont Ferrand, France

^l *Sanofi, 94250 Gentilly, France*

^m *INDS, 94220 Charenton le Pont, France*

ⁿ *Medtronic, 92100 Boulogne-Billancourt, France*

^o *Les entreprises du médicament (Leem), 75017 Paris, France*

^p *INSERM, 75654 Paris, France*

^q *Univ Rennes, CHU Rennes, Inserm, CIC 1414, 35000 Rennes, France*

^r *DRCI, hospices civils de Lyon, France, 69002 Lyon, France*

^s *CHU de Clermont Ferrand, 63003 Clermont Ferrand, France*

^t *Clinsearch, 92245 Malakoff, France*

^u *Astrazeneca, 92400 Courbevoie, France*

^v *Service de pharmacologie clinique, plateforme de recherche clinique de l'Est parisien (URC-CRC-CRB), APHP, hôpital Saint Antoine, 75571 Paris, France*

^w *CNIL, service santé, Paris, 75017 Paris, France*

^x *Lysarc, centre hospitalier de Lyon, 69495 Pierre-Bénite, France*

^y *Roche SAS, 92650 Boulogne-Billancourt, France*

Received 25 October 2018; accepted 5 November 2018

*Corresponding author. Centre d'investigation clinique, hôpital Louis Pradel, 28 avenue du Doyen Lépine, 69500 Bron, France.

E-mail adress: catherine.cornu@chu-lyon.fr (C. Cornu)

* Articles, analyses and proposals from Giens workshops are those of the authors and no not prejudice the proposition of their parent organization.

Summary

Clinical research on human subjects or their data is confronted with conflicting requirements with, on the one hand, the principle of open science (transparency and data sharing), the possibilities offered by big data and the reuse of healthcare or research data, and on the other, changes to the regulatory and legislative framework, including the general data protection regulation (GDPR).

A roundtable was organized in Giens, France in October 2018 to identify problem areas, the need for clarification and streamlining, and to make recommendations to promote clinical research while ensuring a high level of patient protection. After details were given of these developments, the roundtable participants were able to propose recommendations, primarily 1) to clarify: what is considered anonymized data, and what is “public interest” within the meaning of the GDPR; 2) for the French data protection authority (CNIL) to continue preparing reference methodologies to simplify the approval system; 3) to promote the secondary use of data by making it easier to inform patients and obtain broad patient consent, by specifying the circumstances under which their withdrawal and opposition rights apply, so as to limit the risk of bias; 4) to facilitate access to data warehouses by providing technological and methodological aids. The roundtable also recommends increasing discussions between authorities in Europe on research topics, encouraging French authorities to contribute to the preparation of codes of conduct and setting up a voluntary harmonization procedure to coordinate the opinions of data protection authorities, while ensuring that key documents are available in English.

KEYWORDS

Data protection; Regulatory texts; Clinical research; Big data; French data protection authority

Abbreviations

AI: artificial intelligence

CEREES: Expert committee on research, studies and evaluations in the field of health

CNIL: French data protection authority

CNRIPH: National committee for studies involving human participants

DC: data controller

DPO: data protection officer

EC: ethics committee

EKG: electrocardiogram

FAIR: findable, accessible, interoperable, reusable

GDPR: general data protection regulation

HIPAA: Health insurance portability and accountability act

INDS: National institute of health data

MS: Member State (of the European Union)

NIR: directory identification number

PHC: French public health code

PMSI: program for the medicalization of information systems

RM: reference methodology

SNDS: national health data system

SNIRAM: national health insurance cross-scheme information system

WHO: World Health Organization

Introduction

Similarly to many areas of social activity, clinical research is in the midst of a digital revolution through access to massive data and the possibility of reusing digital data.

High-throughput “-omics” and imaging techniques now make it possible to generate big data for each patient, and harness this data using algorithms to stratify patients according to complex biomarker profiles. This is what “personalized” medicine seeks to achieve.

At the same time, the digitization of clinical research data and healthcare data, or health system data, allows the data to be reused in research for a secondary purpose, either for clinical trials or observational studies. These secondary uses include re-analysis, secondary analyses and meta-analyses. It is also possible to reuse cohorts or registries to conduct interventional studies [1–3]. Information from hospital data warehouses can be used to select sites for a clinical trial, to select patients, optimize the study design, or to collect data for a clinical trial or “real-world” study. It will also become increasingly common to use data from connected objects for clinical trials or real-world studies.

In addition, open science policy, pursuant to the findable, accessible, interoperable, reusable (FAIR) principle and for the purpose of improving the reproducibility of study results, promotes the sharing and reuse of clinical research data.

These developments raise multiple issues regarding methods, data protection, de-identification of data and the risk of re-identification, access to and sharing of data, information and re-information of data subjects, and how to obtain consent [4]. The regulatory and legislative framework for these issues is also undergoing substantial changes, particularly with the implementation of the GDPR (Table 1).

These topics were discussed at a roundtable organized in Giens, France in October 2018 to identify problem areas and the need to clarify and streamline, and make recommendations to promote clinical research while ensuring a high level of patient protection. The program covered the new regulatory framework for clinical research (and its implications for information to patients, broad consent and withdrawal of consent), simplification and structuring mechanisms put in place by the French data protection authority (CNIL) and its reference methodologies (RMs), the impact on public and private institutions’ governance, compliance tools, the role of the data protection officer (DPO) [Table 2], the European and international context, data sharing, technical aspects of data security and confidentiality, minimization (anonymization vs. pseudonymization) and the risk of re-identification, the conditions for hosting health data, aggregation of non-downloadable data, sharing and reuse of research data for clinical research (clinical trial data - repositories, security,

conditions for access and secondary analysis, data sharing plan) or of data from the healthcare system, and the use of data from connected objects.

Context

Regulatory context

Regulation (EU) 2016/679 of the European parliament and of the council of April 27, 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, repealing Directive 95/46/EC (GDPR) entered into force on May 25, 2018 [5]. The objectives of the GDPR are to harmonize data protection rules in Europe, to reinforce and modernize individuals' rights with regard to their personal data, to hold those involved accountable, and to facilitate data flows. Figure 1 recalls the main principles of data processing.

At the same time, the regulations governing clinical research are evolving with the forthcoming entry into force of regulation 2014/536 of the European parliament and of the Council on clinical trials involving medicinal products for human use which repeals Directive 2001/20/EC [6]. In the field of medical devices, a new regulation was adopted by the Parliament and the Council in May 2017 (Regulation 2017/746) [7] to harmonize the conditions for the approval and conduct of clinical investigations concerning medical devices and their access to the market.

Secondary use of the data, outside the clinical trial protocol, should comply with applicable data protection laws, including the GDPR and national legislation. However, Article 28 of Regulation 2014/536 explicitly foresees the possibility for the sponsor to ask clinical trial participants, when they give their consent to participate in the trial, to agree to their data being reused "*outside the clinical trial protocol exclusively for scientific purposes*" [6]. This broad consent, which can be withdrawn at any time and which is used as a basis for secondary use of the data, therefore applies to clinical trials on medicines. However, the same article makes it clear that scientific research that harnesses data outside the clinical trial protocol is conducted in accordance with applicable data protection laws. The relationship between clinical trial regulations and the GDPR is currently under discussion in Europe.

French legislation has opened access to health data, such as health insurance data and data from healthcare facility activities, thus facilitating the conduct of research and studies in the fields of health, care and social support.

Data protection

Special provisions on the processing of health data coexist in France with the GDPR. One chapter of the modified French data protection act is devoted to research, studies and evaluations in the field of health (chapter IX, section 2) [8]. Certain conditions must be met, such as the study being for the purpose of public interest, the opinion of a preliminary committee or individual information given to data subjects. Procedures for requesting approvals are described in Figure 2.

To lighten the administrative burden of study sponsors, simplified procedures have been put in place since 2006 through reference methodologies (RMs). These RMs, which are now part of the reference systems, require a high level of confidentiality and protection for individuals. They exempt the data controllers from submitting an application for approval to the French data protection authority (CNIL). Demonstration of compliance with a reference methodology replaces the approval system.

Lastly, the French data protection act applies to data controllers located in France (even if they do not process data of individuals in France) and to data controllers located outside France if they process the health data of individuals who live in France (Table 3).

Data minimization: anonymization versus pseudonymization

The GDPR applies to any personal data relating to an identified or identifiable natural person. Article 2 of Act 78-17 of January 6, 1978, known as the French data protection act, clarifies that “in order to determine whether a person is identifiable, all the means available to the data controller or any other person that enable the person to be identified must be considered”. Therefore, by definition, personal data and anonymous data are opposed to each other. If there is a link allowing the person to be re-identified, the data is to be considered as personal. The data will be anonymous only when it does not allow the person to be re-identified, including by cross-checking with a new database. This anonymous data is not affected by the regulations on the processing of personal data.

Working on anonymous data would enable researchers to avoid the provisions of the GDPR, such as informing participants, maintaining an adequate level of security, and formalizing any data processing.

America's Health insurance portability and accountability act (HIPAA)¹ proposes to “de-identify” personal data so that they can be shared without restrictions, by deleting a list of data defined as identifiers, such as names, zip codes, dates of birth, etc.

The European regulatory framework favors the principle of data minimization (Article 5.1.c of the GDPR) in order to process only the data that is strictly necessary, without prejudice to the actual means used.

The possibility of identifying individuals is interpreted in Europe in terms of the means that may reasonably be used, taking into account all the objective factors such as the cost and time required for identification and the technologies available at the time of processing but also in terms of how they evolve. The concept of anonymity must be reasonably robust over time.

However, anonymization methods are complex and their robustness over time is difficult to assess [9-14]². Therefore, any health research data should be considered non-anonymous personal data until proven otherwise.

¹ HIPAA: Items to remove to de-identify health data: (A) names, (B) geographic subdivisions smaller than a state, (C) all elements of dates (except year) for dates that are directly related to an individual, including birth date, (D) telephone numbers, (L) vehicle identifiers, (E) fax numbers, (M) device identifiers and serial numbers, (F) email addresses, (N) Web Universal Resource Locators (URLs), (G) social security numbers, (O) Internet Protocol (IP) addresses, (H) medical record numbers, (P) biometric identifiers, including finger and voice prints, (I) health plan beneficiary numbers, (Q) full-face photographs and any comparable images, (J) account numbers, (R) any other unique identifying number, characteristic, or code, (K) certificate/license numbers.

² Can we make data anonymous and how can we facilitate sharing for research?

Several anonymous open data dissemination experiments in the United States (GIC 2002, AOL 2006, Netflix 2010) [9,10] have led to individuals being re-identified. In 2016, in Australia, the dissemination of national health system data was interrupted after it was discovered that people could be re-identified [11,12]. In such a sensitive area like healthcare, free access to data therefore seems to be difficult to uphold, given the privacy risks involved. In its analysis of anonymization techniques, the Article 29 [9] group considers that no technique guarantees anonymization entirely without significantly deteriorating the data. True anonymization is a complex process using a combination of generalization techniques (k-anonymity, l-diversity) and randomization techniques (addition of noise, differential privacy, etc.) [13]. These complex techniques reduce the amount of information in the data and the ability to reuse them for research. An alternative to these robust anonymization techniques appears to be reinforced minimization, which consists of removing the main identifiers and semi-identifiers (list of HIPAA items in the USA), replacing all the dates by timeframes and generalizing certain variables when possible. A database prepared using this technique should be able to be shared with the scientific community through a third-party platform that guarantees data access and security using processes such as those described by Ohmann et al. [3, 14]. In order to make transparency requirements compatible with personal data protection requirements, in addition to introducing sharing tools, codes of conduct should be put in place to help share health data for scientific research, in accordance with recitals 26, 33 and 50, and pursuant to Articles 14.5, 40 and 89 of the GDPR.

Access to medical administrative databases and hospital data warehouses

National health data system (SNDS)

French health system modernization legislation adopted in January 2016 ratified the creation of a valuable medical administrative database: the SNDS [15]. This is a pseudonymized database containing data on the use of care in private practices and healthcare facilities for the entire French population, as well as data from the hospital program for the medicalization of information systems (PMSI), and from the causes of death registry, and it will soon receive data on disabilities and private health insurance data. These data are henceforth widely available, the quid pro quo being more stringent conditions ensuring data confidentiality and integrity and the traceability of access, as well as feedback to citizens through transparency about the research conducted using these data and the ensuing results.

Regarding access procedures, simplified procedures (RMs) were already introduced in July 2018. If processing does not comply with the RM requirements, approval must be requested from the French data protection authority (Figure 2).

But the SNDS is above all a formidable stepping stone that will be especially valuable when combined with other sources of clinical, environmental and sociological data. The relaxing of the conditions of use of the directory identification number (NIR [social security number]) should facilitate such combinations, which will have to be processed in very secure environments. In this respect, expectations are high for the “health data hub” prefiguration mission [16].

Changes to medical records

The use of the computerized medical record, designed in university laboratories in the early 1980s and originally intended to manage healthcare data and codify diagnoses and procedures, has changed as it has become increasingly standardized since the 1990s. It is now primarily a tool for monitoring medical activities and billing. As a result, it is often a digital copy of what was the paper file with no added value for research. The data are difficult to harness because of the multiple applications involved, their low interoperability, and the lack of structure for much of the data. The advent of technologies to handle large amounts of data and the increasing use of standardized data formats now offer the possibility of building health data warehouses on the scale of a hospital, or

even a region, to aggregate all of a patient's data. The new full-text search and indexing technologies, through semantic enrichment via medical ontologies, make it possible to index and query unstructured textual documents such as medical observations and examination or hospitalization reports, which are a major source of information. These hospital clinical data supplement the data from the national health insurance cross-scheme information system (SNIIRAM), which provide valuable information on the use of care services and the path of patients who have received hospital care (for those patients who did not receive hospital care, diagnoses are not available in this data base).

On the scale of a hospital or region, access to these data must be controlled (ethics, regulations, methodology, information to patients, transparency about processing). The French data protection authority, which issues the necessary approvals, attaches particular importance to data security, information to patients and the governance of these warehouses. The solution implemented at the largest university hospital in France (Paris) and the eHOP project at Rennes university hospital include the creation of a "trusted third-party" structure responsible for analyzing requests, processing them and providing patient data needed for a project, in compliance with ethical standards and regulations.

In the near future, these data will be supplemented by data provided by patients (self-quantified data), genetic data (*-omics*) and data collected routinely (data from monitors, ventilators, electrocardiograms [ECG]). Fast-growing artificial intelligence (AI) technologies will be able to take full advantage of these data, particularly by mining unstructured data in data warehouses.

Data from connected objects

Clinical trials and real-world studies can also be performed using data from connected objects and remote monitoring tools. There are more and more disorders allowing this type of monitoring (diabetes, heart failure, cardiac arrhythmia, etc.). RMs cater for these data collection methods but many remote monitoring tools were introduced without foreseeing the recovery of data for clinical research, or even the recovery of data to improve the algorithmic efficiency. This leads to the essential question of how to inform patients and record their non-objection, and issues of security, confidentiality and quality of data from connected objects used in the research.

Real-world studies using healthcare databases, cohorts and clinical trials

The possibility of conducting studies using medical administrative databases is already transforming the landscape of clinical research. The Constances cohort, representative of the French population, includes 200,000 volunteers aged 18 to 69, and combines: 1) visits to one of the participating health insurance funds' health examination centers, to have a medical and paraclinical examination and lab tests and, for those aged 45 and over, cognitive and physical tests; 2) a biobank; 3) the use of data from medical administrative databases (national pension fund, health insurance fund, hospitals and healthcare providers); and 4) questionnaires on various topics related to the health status, regularly completed by volunteers (Table 4) [17,18].

Similarly, the iVasc project (co-funded by public and private funds), includes a cohort of patients (FRENCHIE) hospitalized for myocardial infarction, monitored thanks to the medical administrative databases, and also makes it possible to conduct interventional studies nested into the cohort, two of which are already planned and funded - on dental health and sleep disorders.

The scientific and ethical necessity of data sharing

The GDPR must be enforced within the broader framework of the open science movement that is revolutionizing the standards hitherto accepted by scientists by imposing three key scientific values: transparency, openness and reproducibility. Data sharing has an ethical justification. In participating in the studies, the individuals take risks in exchange for uncertain benefits. The result is an implicit agreement whereby the data from these studies must bring the greatest possible benefit to society as a whole. To maximize these benefits, as of January 2019, the International committee of medical journal editors (ICMJE) expects: 1) authors reporting clinical trial results to detail their intentions about sharing de-identified data and associated metadata [19] and 2) a data sharing plan to be prepared and recorded before the start of the clinical trial, at the same time as the trial is recorded on a ClinicalTrials.gov-type registry [20-22]. Several journals, including the BMJ and the PloS group journals, preceded this movement, encouraging authors to provide their data for several years. Nevertheless, the GDPR can complicate compliance with these transparency requirements.

How to manage the conflict between protecting personal data and transparency about research/data sharing

There is indeed an apparent conflict between the need to strengthen the protection of personal data and the equally ethical need to share healthcare or research data to improve knowledge and, ultimately, patient care. Some provisions in the GDPR on patient information and consent regarding secondary use could jeopardize this progress. These provisions could undermine the scientific validity of data reused for research purposes. In particular (except for the use of anonymous data that are not affected by the GDPR), the possibility of not consenting to secondary use, of consenting only to restricted use (for a given institution or disorder), and more importantly, the possibility of withdrawing consent for secondary use could increase the risk of bias if many patients exercise this right - especially if the patient's satisfaction during hospitalization, or during a study, affects their decision to withdraw consent for secondary use.

A survey of 771 clinical trial participants in the United States showed that 93% of respondents were likely to allow their data to be shared with university scientists, and 82% with

scientists at for-profit companiesⁱ [23] - a better acceptance rate than for the sharing of hospital data or biological samples.

Impact of GDPR on clinical research

Reference to the GDPR is therefore essential for all aspects of clinical research, from information to patients, which must now refer to the GDPR, and the type of information that the regulation advocates, to the publication of the results, where it can come into conflict with transparency and data sharing requirements. To mitigate these problems, we propose the following recommendations.

Recommendations

These recommendations fall into three categories: the exercise of individuals' rights, streamlining administrative procedures, and a more open relationship with the international environment.

Recommendations concerning the exercise of individuals' rights concerning their data

GDPR and information to patients

The entry into effect of the GDPR reinforces the obligation to inform patients, who are, like any other citizen, bombarded with messages from service providers, businesses, etc. about the GDPR update, which may prevent the information about their participation in the study from getting across. To avoid this, the roundtable recommends providing a **template for patient information form content** (template proposed by the French national committee for studies involving human participants [CNRIPH]). A proposed guideline specifying what information to give to which

patients (new patients, patients already included, patients who have completed a study), when, and by whom was disseminated on May 25, 2018 by the CNRIPH, and is currently being reviewed.

Exercise of patients' rights

The right to erasure and the right to opposition may conflict with the methodological requirement to conduct exhaustive data analysis for the primary use or reuse of data. Indeed, the missing data undermines the validity of the estimated treatment effect in randomized controlled trials. To avoid bias, it is advisable to do an intention-to-treat analysis, including all randomized patients in the group to which they were randomized. Missing data rates need to be specified, and specific methods for taking into account missing data should be used and explained, but these methods do not completely eliminate biases related to missing data [24,25]. This methodological requirement is also included in Regulation 2014/536 of the European Parliament and of the Council on clinical trials involving medicinal products for human use which states (Art. 28.3) that “withdrawal of informed consent does not affect the results of activities already carried out, such as the storage and use of data obtained based on the informed consent before it is withdrawn”. We recommend **clarifying under which circumstances this right of withdrawal applies to secondary use** (safety and efficacy data), how to inform patients about the negative effects on the study conclusions, and under which conditions patients should be allowed to knowingly and differentially refuse to continue the study treatment, the monitoring foreseen by the protocol, the collection of data during the usual monitoring for the purposes of the study, or even the erasure of their data. The French data protection authority/CNRIPH are expected to propose additional information for patients.

The conditions of the right of opposition also raise questions: patients have the right to object to their data being processed, but the practical aspects of exercising this right are not simple. **It needs to be specified and clarified through whom these rights are exercised:** firstly through the investigator (the only person the patient knows and who knows his or her identity) and/or the sponsor's DPO who can be aware of the identity of the person who wishes to contact him or her; the DPO is bound by an obligation of confidentiality or professional secrecy in the performance of his or her duties.

The right to portability is included in the GDPR, where the legal basis is the consent or contract, to ensure patients' medical data is available to them for the continuity of their medical care when they change their place of care. This provision has little relevance for data collected specifically for research. The explanation given to patients about this right becomes unintelligible

unless, for a given study, it details how this right might apply. **An exception to this right should be made for study data** that are not useful for care.

Reuse of data

- The RMs define the possibility of secondary use, with information to the patient. This issue of informing the patient can arise prospectively, when planning a study, to anticipate future uses, or when a database created earlier is reused. The recommendation is to ensure that **the initial information given to the patient takes into consideration subsequent studies**. In order to minimize its limitations, while preserving the confidentiality of the data, **the initial information on reuse should focus on how to access the data rather than on the topic of subsequent studies**: who can access the data? What data (what kind of anonymization)? In what ways: controlled access (via a request analyzed by a scientific committee), possibility of aggregating non-downloadable data, sharing after anonymization? Which repository will be used, and what is its governance?

Thus, **RM-004 could be updated with another change to the individual information**:
“When the study is based on clear criteria that the data subject has been made aware of in advance regarding the use of his/her data according to the purposes, data access methods and persons who have access, and the data protection measures applied.”

- Promoting the participation of **patient representatives in the governance of repositories** allowing secondary use of data.
- **Promoting incentives for data sharing**: it seems important for study value indicators to include this sharing, the secondary publications resulting from this sharing, the number of accesses to a database, etc. The GDPR maintains the requirement for patients to be individually informed about the secondary use of their personal data, but there are possibilities of exceptions, especially *if the provision of such information proves impossible or would require disproportionate efforts, or if it would seriously compromise the achievement of processing objectives*. Insofar as failing to provide information to patients makes the research ineligible for the reference methodologies (RM), these situations are analyzed on a case by case basis by the French data protection authority, based on the file provided by the applicant, when applying for approval (Chapter IX of the data protection

act). The roundtable's recommendation is for the French Data Protection Authority to make available: 1) a guideline specifying the **criteria for granting these exceptions**, and 2) **examples of what can be considered as a "disproportionate effort"** (European guidance on "transparency" [22]).

Recommendations concerning the fluidity of administrative procedures

- It is important to **continue developing reference methodologies** and **avoid limiting them to specific entities** (such as RM 005 which applies to health facilities, and *de facto* excludes universities and other public research bodies).
- It is important to propose a framework for **defining public interest**, a necessary condition for data processing for research purposes in France.
- Datasheets must be created to **help researchers who want to access the various data warehouses**, especially those of the SNDS, including an inventory of existing databases and methods of access.

In particular, it is necessary to **clarify what is considered anonymized or pseudonymized data**, propose validated anonymization solutions, and IT tools and training for data managers to anonymize data.

- To facilitate matching and make linking possible, removing the contradiction stemming from the fact that **the use of the NIR is now possible** without a decree of the *Conseil d'État*, unless **it comes from a medical file**.

Recommendation related to the single GDPR annex for the "single agreement": need for a contract between the data controller and the processor

The GDPR has redefined the responsibilities and obligations of all those involved in the processing of personal data. The data controller ("DC") is the person who determines, alone or with others, the purpose and method of processing. The processor is the entity that processes personal

data on behalf of the DC. The processor has new obligations under the GDPR and may now be held liable in case of default.

The single agreement was introduced by the health system modernization law (“LMSS”) of January 26, 2016, with the aim of improving timelines for setting up clinical trials in French hospitals, which increases France's attractiveness for international clinical studies [26]. The single agreement model to be used was set by the decree of November 16, 2016, provided for in Article R. 1121-4 of the French public health code (PHC), prior to the GDPR. Its scope is type 1 and 2 studies involving human participants conducted for commercial purposes in health facilities, care homes and health centers.

In the context of the single agreement, under the GDPR the DC is the sponsor of a commercial study involving human participants. The processor is the health facility, care home or health center that signed the single agreement.

Article 28 of the GDPR defines the contractual framework governing the relationship between the DC and the processor. According to this article, the DC must make a contractual agreement with the processor to ensure that the latter meets all the conditions and offers all the appropriate guarantees to protect personal data. Article 28.3 of the GDPR gives a non-exhaustive list of the clauses to be included in the contract.

Article 28.3 of the GDPR states as follows: “Processing by a processor shall be governed by a contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller and that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller”.

Moreover, the reference methodology RM 001 adopted by the French Data Protection Authority, which regulates the processing of health data of public interest, carried out in the context of research requiring consent, reiterates the obligation to establish a contract between the DC and the processor.

It is therefore necessary to introduce a subcontracting clause into the single agreement to ensure that it complies with the GDPR. An update could be made by modifying the single agreement template by order of the Ministry of Health, notwithstanding the ongoing discussions on several other issues related to the implementation of the single agreement.

Improving technical aids

- A section is expected to be introduced into the RMs on **how to send individual data to peer reviewers** of scientific journals to provide secure sites corresponding to these specifications for researchers.
- The INDS directory, which records studies conducted using the SNDS, should meet the **World health organization's (WHO) requirements for registries** [27], which are mandatory by law and for publishing. This would avoid authors having to register their studies a second time in a compliant registry approved by biomedical journal editors. These registries must meet several criteria and include the minimum 20 items [28].
- Some studies require clinical study patients' details to be registered, for example to reimburse costs. However, the French data protection authority recommends complete separation between databases containing personal health data and those containing personal data for reimbursement. This total separation also concerns university hospital sponsors. University hospitals could declare as processors the administrative services that have patients' contact details for reimbursement purposes. **It must be specified whether this organization can fit into the framework of RMs.**
- The use of SNDS data is governed by a clear regulatory and technical framework. These requirements also apply to studies or data processed from these studies (“extensions”), which complicates the cascading reuse of data from the SNDS. Legislation, as it stands, may be interpreted restrictively, whereby the SNDS may be used for studies, research or evaluations, which would prohibit the creation of data warehouses matched with the SNDS. In addition, these restrictions are no longer relevant when SNIIRAM data used in the extensions become minimal. It would be helpful to give a **precise definition of what is an extension.**

Fostering international cooperation

- Harmonization of data protection regulations and procedures across Europe concerning health research needs to be better documented. For example, it is difficult at present to ascertain to what extent the GDPR and the obligations of multinational clinical trial sponsors have been integrated in each European country. Therefore, an assessment is needed in the form of a **comparative European analysis of national legislations** and associated

health research data protection requirements. This proposal could be included on the agenda of Horizon 2020.

- If there are significant differences between Member States, a **voluntary harmonization procedure could be proposed. Its purpose would be to coordinate the opinions of the data protection authorities** of the different Member States. Regarding obligations to Ethics Committees, in order to facilitate international projects, **foreign sponsors should be told** how to proceed and make the documents available. The CNRIPH is committed to working on this.
- Concerning the GDPR, associations and other representative bodies may prepare codes of conduct to clarify how the GDPR applies to certain sectors, including scientific research. It will therefore be necessary to **increase the involvement of French stakeholders and authorities in the preparation of codes of conduct** concerning data processing in the context of clinical trials, in particular for multinational processing. Such a code is currently being prepared by various European partners [29].
- The French data protection authority has put in place RMs, which have greatly simplified the clinical study data protection process. These tools and procedures have often been translated into English for internal use by different companies. An **“official” translation of the RMs** should be provided by the French data protection authority, and they should be more easily available to foreign sponsors.
- Lastly, the **“health data host” certification** in France requires a specifically French procedure, in connection with a reference system based on compliance with international standards (ISO 27001, 20000, 27018 and 27017). It would be helpful to move towards a more interoperable system so as not to isolate France’s health data.

Conclusions

Clinical research involving human participants or their data appears torn between, on the one side, the requirement for transparency and data sharing according to the principle of open science, the possibilities offered by big data (-omics), or by the possibility to reuse data (hospital data, health databases, study data - cohorts and registries), and on the other, the ill-defined constraints of the GDPR.

This is a hot topic for all concerned. France must avoid isolating itself in a barely interoperable regulatory system and can be capable, with a few adjustments, of protecting its citizens' data while streamlining clinical research. The main recommendations resulting from this roundtable are first of all to clarify what are considered as anonymized data (not concerned by the GDPR) and pseudonymized data (personal data, covered by the GDPR). A clarification of what is considered to be of public interest would also be welcome, as it is a condition for the use of data for research purposes in France. It is also important for the French Data Protection Authority to continue to develop reference methodologies to simplify the approval system and avoid limiting them to specific entities.

In order to promote the secondary use of data, it is important for the patient's initial consent to encompass subsequent research, and describe how the data will be accessed rather than restricting access to subsequent research on a limited topic. It should be clarified under what circumstances the right of withdrawal applies to secondary use, as it may introduce bias into secondary analyses. Researchers who want to access the different repositories and warehouses to reuse data must have access to a dedicated infrastructure providing technological and methodological aids.

Lastly, we must avoid isolating France from the European and international communities, by using a European benchmark on laws and regulations affecting the protection of personal data in clinical studies, by discussing research topics with other European authorities, by encouraging the French authorities to contribute to codes of conduct, by proposing a voluntary harmonization procedure to coordinate the opinions of the data protection authorities, and by ensuring that the main documents are available in English.

Disclosure of interest

The authors have no conflicts of interest to declare.

References

- [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. *Therapie*. 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> [Accessed 23 November 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> [Accessed 23 November 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> [Accessed 23 November 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> [Accessed 23 November 2018]
- [9] Commission européenne. Article 29 Newsroom - News overview - European Commission. Mai 2018. <http://ec.europa.eu/newsroom/article29/news-overview.cfm> [Accessed 23 November 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> [Accessed 23 November 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> [Accessed 23 November 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>. [Accessed 23 November 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 [Accessed 23 November 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/> [Accessed 23 November 2018]

[15] Système national des données de santé. Septembre 2018. <https://www.snds.gouv.fr/SNDS/Accueil> [Accessed 23 November 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> [Accessed 23 November 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>. [Accessed 23 November 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] Hrynaskiewicz 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.

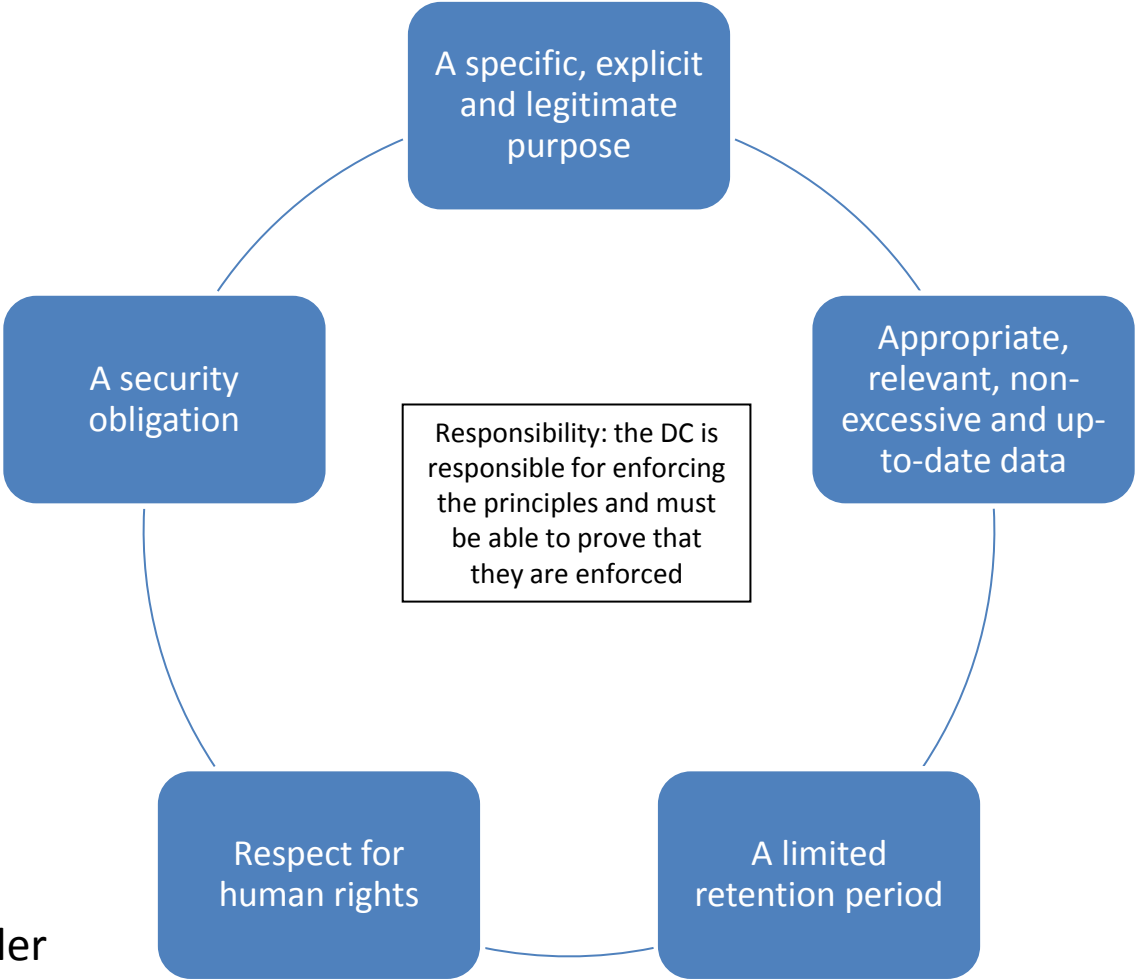
- [22] Bonini S, Eichler HG, Wathion N, Rasi G. Transparency and the European medicines agency--sharing of clinical trial data. *N Engl J Med*. 2014 Dec 25;371(26):2452-5. doi: 10.1056/NEJMp1409464.
- [23] Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of 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 [Accessed 23 November 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^e 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> [Accessed 23 November 2018]
- [28] World health organization. International clinical trials registry platform. WHO data set. 2018. <http://www.who.int/ictrp/network/trds/en/index.html> [Accessed 23 November 2018]
- [29] BBMRI-ERIC. GPRD code of conduct. 2018. <http://www.bbmri-eric.eu/BBMRI-ERIC/gdpr-code-of-conduct> [Accessed 23 November 2018]

Table 3. The reference methodologies of the French data protection authority (CNIL).

Types of research	Research involving human participants (RIHP) <i>Opinion from EC in all cases</i>			Research not involving human participants, health evaluation or study (RNIHP)	
	Category 1 Interventional studies	Category 2 Interventional studies involving minimal risks and constraints	Category 3 Non-interventional studies	Research organized and conducted on humans with purposes other than those of an RIHP	Research involving data or samples collected in another context (reuse of data)
Scope of RMs	RM-001 (written or express consent required)		RM-003 (information and non-objection)	RM-004 (information and non-objection – changes to individual information) <i>No opinion from CEREES.</i>	
			RM-001 if genetic characteristics are studied (as consent is required)		
If non-compliance with RMs*	Favorable opinion from EC + approval from French Data Protection Authority (+ possibly referred to INDS to consider public interest)			INDS (single secretariat + public interest) + favorable opinion from CEREES + approval from French Data Protection Authority (+ possibly referred to INDS to consider public interest)	
	*In particular regarding: information provided to the person, type of data processed, recipients of direct or indirect identifiers, high residual risk etc.				

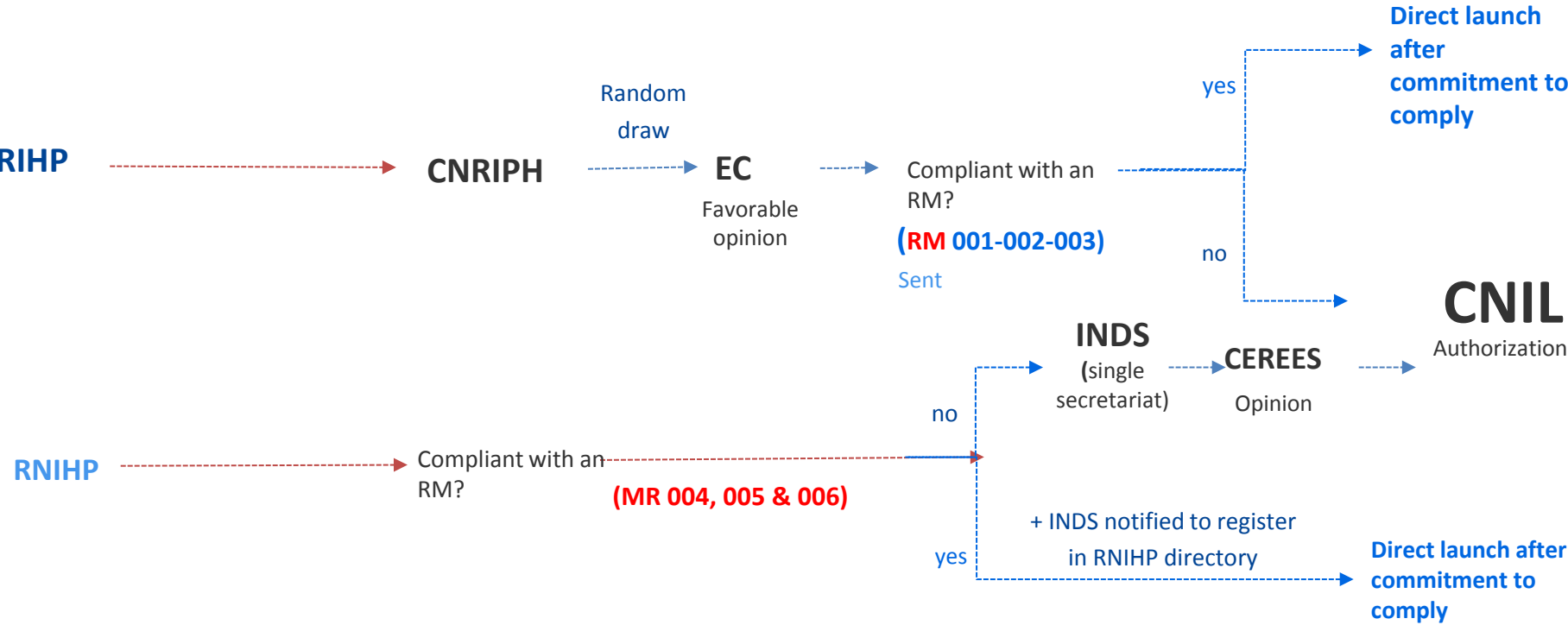
CEREES: Expert committee on research, studies and evaluations in the field of health; CNIL: French data protection authority; EC: ethics committee; INDS: National institute of health data; RIHP: research involving human participants

Figure 1. Basic principles of data protection



DC: data controller

Figure 2. Regulatory measures for research involving human participants or otherwise.



TYPE OF RESEARCH

- RIHP** *Research involving human participants*
- RNIHP** *Research not involving human participants*

INSTITUTIONS

- CNRIPH** *National committee for research involving human participants*
- INDS** *National institute of health data*
- CEREEES** *Expert committee on research, studies and evaluations in the field of health*
- EC** *Ethics committee*
- CNIL** *French data protection authority*

SIMPLIFICATION

- RM** *Reference methodology*

Table 1. Stages of clinical research impacted by the GDPR

<i>Point of impact</i>	<i>Explanation / types of database</i>
Site selection, investigator selection	A site/investigator can be selected based on its past publications, its activity and its patient population
Agreement between sponsor and investigational site	Agreement needed between data controller (DC) and processor
Patient selection	Patients can be selected from healthcare databases: PMSI, data warehouses
Informed consent	GDPR information updated, broad consent, re-consent, evolving consent, dynamic consent, e-consent
Cohort/registry data Data from national databases	Specific studies or clinical trials using these databases
Electronic health records	Data on care is inserted into research databases and clinical trial CRFs
Electronic data capture	Connected objects, e-questionnaires
Data sharing (FAIR)	With medical journal reviewers, and made available to the community
Reuse of data	Access to existing databases

CRF: case report form; FAIR: findable, accessible, interoperable, reusable; GDPR: general data protection regulation; PMSI: program for the medicalization of information systems.

Table 2. Role of the data protection officer (DPO)

Each public or private organization must organize the governance of its processing. Those concerned must be trained and be familiar with data protection procedures. The DPO plays a key part in compliance, ensuring that it is upheld. Different tools are used (data processing register, procedures, etc.), a key one being the “privacy impact assessment”, which is mandatory for healthcare research. It is an exercise in describing the project, its legal framework, the security precautions put in place, the information given to concerned person and the risk analysis. This documents the processing done and makes it possible to evaluate the risks.

All these items allow the organization to document its compliance and fully comply with the principle of “accountability” in the general data protection regulation (GDPR) and conveyed in reference systems such as the reference methodologies.

Table 3. The reference methodologies of the French data protection authority (CNIL).

In July 2018, three reference methodologies (RMs) were updated from existing RMs (RM 001, RM 003 and RM 004) and two others were created ex nihilo (RM 005, RM 006). The two latest RMs concern access to hospital activity data (program for the medicalization of information systems [PMSI]) and apply to healthcare facilities and federations, as well as to healthcare product manufacturers.

RM 001 and RM 003 are for research involving human participants, and RM 004 is for research not involving human participants (research on data, or prospective studies that are not organized and conducted on humans to improve biological or medical knowledge e.g. studies aimed at studying professional practices, teaching modalities ...). The main principles of the RMs are the processing of health data that can identify data subjects indirectly (coded or pseudonymized data) by the data controller, the individual information given to data subjects and an analysis of the impact on their privacy.

The new elements introduced in July 2018 concern in particular the obligation (except in specific cases provided for by the regulations) for the data controller to appoint a DPO, the possibility for a subcontractor of the data controller (processor) to process data directly identifying the data subjects (reimbursement of costs, payments, text message sent to complete a questionnaire), the possibility of processing the Department where the data subject lives, sharing data with independent experts or researchers responsible for re-analyzing the data, particularly at the request of scientific journal editors (technical solution only allowing the data to be read, without being able to extract personal data). Data from extensions of the National health data system (SNDS) must be used in accordance with the requirements of the SNDS provided for in the French public health code.

Lastly, for RM 004, which concerns processing requiring the reuse of data, changes were introduced concerning individual information given to data subjects; it is not necessary to individually inform a data subject again when the latter already has this information (for example: several studies, handled by the same data controller and whose data categories have the same purposes and where the recipients are the same) or cases where, when biological data and/or samples are collected, the data subjects are directed to a specific information system that they can refer to before each future treatment (for example a website). The list of studies conducted according to an RM must be registered by the data controller in the public directory of the National institute of health data (INDS).

Table 4. Conditions for secondary use of the Constances cohort data [17].

Due to the regulatory and technical complexity of data collection circuits, creating the Constances cohort required many months of reflection and exchanges with the supervisory authorities and institutional partners to guarantee the mobilization and protection of personal data. To achieve the goals of the cohort (i.e. to make available to research teams a selection of data necessary for their research, in a very secure environment), the researchers have to submit a project that will be validated by the international scientific council and those responsible for the governance of Constances. The objectives of this research may relate to public health issues, epidemiological research, but also clinical research by offering, for example, the possibility for a study's design to create a real-world control arm selected based on factors matched to the population included in the interventional arm(s). Under these conditions, depending on the type of project and the data mobilized, before being able to work on the specific secure space available to them, the researchers have to comply with the regulatory framework of a reference methodology or solicit the opinion (or approval) of any relevant authorities (ethics committee [EC], expert committee on research, studies and evaluations in the field of health [CEREES], the French data protection authority [CNIL], the French national agency for medicines and health products safety [ANSM], and the National council for statistical information [CNIS]). The volunteers can see all the projects accepted on the website www.constances.fr, they can subscribe to the newsletter to be directly notified and they receive the cohort study journal each year, which gives general updates and describes the type of data used for each of the projects. Each volunteer can then exercise his/her right to object to this data being used for one or more specific projects.