



Overview of the European General Data Protection Regulation (GDPR) impact on medical writing for clinical trials

Tatiana Revenco¹, Gregory Collet^{1,2}

doi: 10.56012/ivhe5802

- 1 MyData-TRUST, Mons, Belgium.
- 2 Centre for Research in Cognition and Neurosciences, Université libre de Bruxelles, Brussels, Belgium.

 Tatiana Revenco
ORCID: 0000-0001-6467-4598

 Gregory Collet
ORCID: 0000-0002-7378-0074

Correspondence to:

Tatiana Revenco

t.x.revenco@mydata-trust.com

Abstract

The European General Data Protection Regulation 2016/ 679 (GDPR) aims to ensure the security and privacy of individuals in the European Union (EU). Companies located within and outside of the EU must comply with GDPR when processing personal data of EU citizens.

Medical writing includes the development of documents related to clinical research. To develop those documents, medical writers have access to personal data, including health information considered as sensitive data.

Therefore, medical writing falls within the purview of GDPR and must comply with its requirements.

This article is an overview of the impact of GDPR on medical writing including security measures such as anonymisation, pseudonymisation, and data minimisation techniques. It also provides an overview of the technical and organisational actions in the framework of medical writing to guarantee respect of data subjects' rights and freedoms.



Introduction to the GDPR

The European General Data Protection Regulation 2016/ 679 (GDPR) became effective in May 2018 and aims to harmonise data protection laws across EU member states.¹ The goal of GDPR is to render individuals control over their personal data, and to enhance security measures, including information technology (IT) for data protection. GDPR defines personal data as “any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person” (Art. 4 GDPR).

Medical writing under the purview of GDPR

In the framework of clinical research, special categories of personal data are processed including demographics (e.g. age, gender, ethnicity, race), health data, and genetic data, some of which are considered as sensitive data and require more security safeguards (Art. 32 GDPR).^{1,2}

Institutions performing clinical trials often engage service providers for medical writing activities. Since medical writers handle personal data, they are considered as data processors and have responsibilities and obligations listed in GDPR Art. 28 (“Where processing is to be carried

out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject”).¹ However, medical writing service providers are often overlooked as data processors, thus lacking the control and implementation of appropriate security measures to protect personal data. Consequently, the risks for data breaches and harm to an individual’s freedoms and rights that might occur in the medical writing framework are underestimated.

It is important to distinguish between medical writing of scientific publications (including articles in scientific journals, abstracts, and presentations for congresses) and medical writing in a clinical study (including case reports, safety reports). Depending on the type of document to be written, writers have access to different types of data in terms of directly identifiable personal data, anonymised or pseudonymised (Table 1).

Article 4 (5) of GDPR defines pseudonymisation as “the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person”.¹

Recital 26 of GDPR defines anonymised data as “information which does not relate to an identified

or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable”.¹

Consequently, GDPR compliance requirements may differ for different medical writing tasks.

Publications writing

In case of most scientific publications (e.g. systematic reviews and meta-analysis) where writers use aggregated data and group statistics, the risk of re-identification of individuals is almost null as the data may be considered as anonymised and do not fall under the scope of GDPR. However, a thorough analysis of anonymisation techniques must be done to ensure that the data are truly anonymous.

The EU Data Protection Working Party, an independent European advisory body on data protection and privacy, issued under Article 29 of Directive 95/46/EC an opinion regarding anonymisation techniques.³ The European Medicine Agency (EMA) issued guidance for anonymisation techniques specifically for the publication of clinical data.⁷ For personal data to be considered truly anonymous, three cumulative criteria must be fulfilled (Figure 1).

1. No individualisation or singling out: The identification of an individual must be rendered impossible by any means, neither direct, nor by isolation of some information from datasets or combination of datasets.
2. No correlation or linkability: The correlation of records to an individual or to a group of individuals within a cohort is impossible.

Table 1. Documents developed by medical writers that involve handling personal data of study subjects and requiring GDPR compliance

Documents	Study subject’s personal data
Clinical Study Report	Typically, pseudonymous data, but combination of datasets can single out the individual. Exceptionally, directly identifiable data when sending to authorities.
Statistical outputs	Typically, pooled data, pseudonymous data.
Safety reports	Identifiable data, directly identifiable data might occur exceptionally.
Case reports	Identifiable data, directly identifiable data might occur exceptionally.
Articles for scientific journals	Typically, pseudonymous data, but combination of datasets can single out the individual.

Still, the linkability stays valid in this case.

- No inference: Personal data cannot be inferred to an individual, meaning that the probability to deduce a value of an attribute within values of a set of attributes is very low. However, to achieve full anonymisation in practice is almost impossible, as shown by regenerative models that the success of re-identification in incomplete datasets is high.^{4,5} Therefore, caution should be taken when evaluating whether personal data is truly anonymised.

Writing for clinical trials

Patient data in clinical trials are not truly anonymous. To ensure individuals' privacy during a clinical trial as per Good Clinical Practice (GCP), when a subject is enrolled in a clinical trial, a code (e.g. Subject ID) is attributed to replace the name and surname. Hence the individual cannot be directly identified. This procedure is called pseudonymisation. Importantly, pseudonymised personal data fall under the scope of GDPR.¹

When medical writers have access to the directly or indirectly identifiable personal data of study subjects, they fall under the purview of GDPR.

Pseudonymisation procedures commonly used in the framework of medical writing are data generalisation, data transformation, encryption, and hashing (see example in Table 2). However, these techniques are not completely immune to re-identification attacks.

For personal data to be effectively pseudonymised, four cumulative criteria should be met (Figure 2).³

- Firstly, no individualisation is possible, yet by using additional information (e.g. key to the code), the individual can be singled out.
- Secondly, the key to the code must be kept confidentially by the data exporter (that might have the role of data controller or data processor), and typically at the investigational sites of a clinical trial.
- Thirdly, appropriate safeguards must be put in place to avoid data breaches and render to the exporter control over the personal data.
- Lastly, a thorough analysis must be performed to ensure that it is impossible to single out the

individual even in case of cross-reference, considering the availability of such data.

Thus, when publishing results in scientific journals, it is necessary to remove or replace certain elements that might lead to identification of the individual.^{6,7} Some examples of techniques are:

- Perform double coding of patient code that was initially attributed by the investigational site
- Banding: Replace subjects' ages with age ranges (reasonably calculated)
- Relativity: Replace calendar dates with relative dates, i.e. in relation to study milestones such as inclusion, randomisation... (e.g. "visit 1", "visit 2")⁷
- Generalisation or randomisation:
 - Date of birth replaced by year of birth or derived age
 - Avoid mentioning the country and/ or city of the investigational site. Site information elevated to larger geographic area
 - Avoid mentioning the name or the alphabetical code of the investigational site

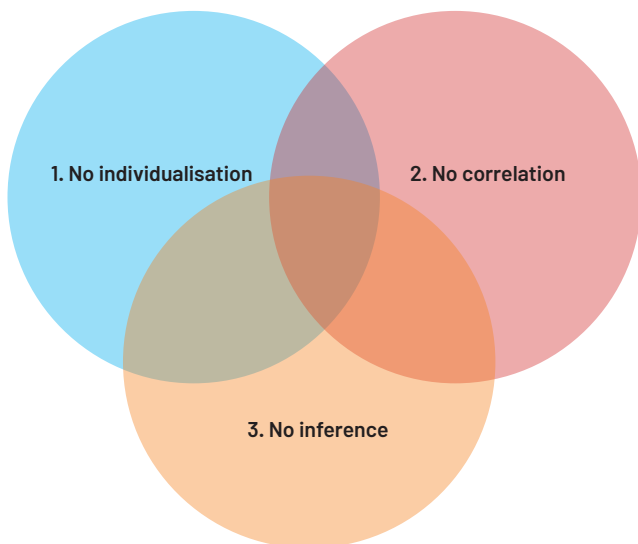


Figure 1. Three cumulative criteria to consider personal data truly anonymised

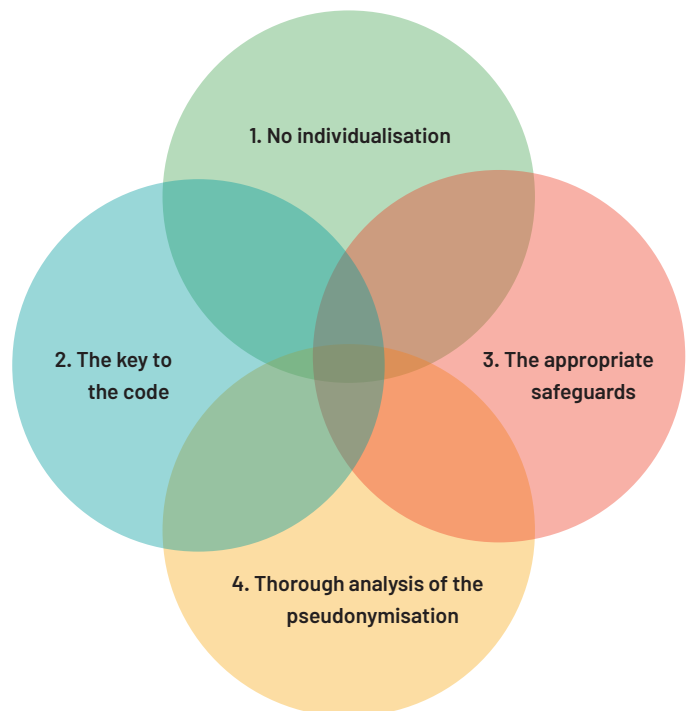


Figure 2. Four cumulative criteria to consider personal data effectively pseudonymised

Table 2. Example of data masking

Name, address, date of birth	Patient's code	Period of treatment	Body Mass Index
	RM54LM286	< 2 years	16
	XD96CV749	> 2 years	18
	SZD95LE206	< 2.5 years	20

- Rare adverse events: preferred term elevated to body system
- Remove or aggregate outliers: e.g. subjects >89 years removed, low frequency groups aggregated

Sharing a minimum amount of personal data to support the scientific findings is in line with the

GDPR principle of data minimisation (Article 5 (c)).

An exception is the publication of individual patient data such as case reports about rare diseases, diagnostic challenges, and treatments of uncommon situations. To ensure patient's privacy, it is compulsory to remove any unnecessary detail and images that can lead to

re-identification of the individual.^{8,9}

When medical writers have access to the identifiable personal data of study subjects, they fall under the purview of GDPR. Access to identifiable data often occurs during writing of the documentation for real-world evidence. One of the data source examples is Council of International Organizations for Medical Science

Table 3. Checklist of some technical, organisational, and IT controls that might help prevent unwanted modifications, loss, or destruction of data, as well as decrease the likelihood and the severity of risks triggered in case of data breaches

Organisational controls

- Set-up data protection policy.
- Set-up data breach procedure describing step-by-step action to be taken to contain the breach.
- Set-up archiving and data destruction procedure.
- Make the personnel aware about the policies and procedures related to data protection.
- Use confidentiality clauses in contracts with processors and freelancers who handle personal data.
- Raise privacy culture.

Logical security controls

- Ensure control and restriction of the access to the sources containing personal data.
- Limit the number of users who may have access to personal data.
- Limit time access to personal data.
- Use robust passwords, secure internet connections, data encryption, installing malicious software on workstations.
- Set-up clear procedures that state how, to whom, by whom, and under what circumstances personal data can be accessed, erased, or sent back to the sponsor.
- Set-up procedures that define traceability methods to track the loggings to the documents containing personal data.
- Encrypt the data.

Physical security controls

- Ensure physical security of servers and platforms of data exchange.
- Ensure physical security of workstations.
- Set-up procedure how to manage paper format containing personal data.
- Set-up policies describing how physical maintenance of hardware is managed.
- Set-up procedures describing actions to take against on-human source of risk.



(CIOMS) forms to report suspect adverse reactions. These forms are completed by the health care professional in free text, hence lacking data protection safeguards.

Also, identifiable data might be accessible, during the writing of a Clinical Study Report (CSR), case reports, safety reports, as well as other documentations for regulatory submissions (Table 1). To draft safety reports, medical writers might use sources such as clinical and safety databases, patient registries, but also patient-generated data through mobile devices, apps, patient-reported outcomes (e.g. eDiaries) and electronic health reports. Moreover, the information contained in the study reports should match with the clinical trial data. This step involves quality control

Medical writers are considered as data processors and play an important role in data protection.

check-up by additional members of the medical writing team and statisticians and increases the risk for data security.

Conclusions

To summarise, it is important to set-up a clear methodology of data pseudonymisation, anonymisation, and data minimisation to ensure the privacy of subjects participating in the clinical trials. Technical, organisational, and IT safeguards must be adopted by the medical writing service providers (Table 3). Writers must respect the data minimisation principle of GDPR by providing only the minimum data necessary to meet the objective of the scientific or regulatory document being written.

In conclusion, GDPR states that data con-

trollers and processors must put in place security measures and ensure privacy by design and by default. However, the regulation does not provide clear instructions as to what those measures are and how they should be implemented. Medical writers are considered to be data processors and play an important role in data protection. Each organisation that provides medical writing services should adopt the necessary measures according to its budget and size in order to comply with data protection regulations. Medical writers as freelancers are also subject to the above-mentioned requirements. With the new EU Clinical Trials Regulation 536/2014, more and more emphasis is set on public disclosure of documents which increases this need for security. An infallible system does not exist, and data breaches occur daily. Therefore, it is important to propagate privacy culture and raise data protection



awareness within the medical writers' community to ensure that subject's rights and freedoms are not compromised.

Acknowledgements

The authors would like to thank Andrea Karambelas, PhD, for proofreading.

Disclosures and conflicts of interest

Both authors are employees of MyData-TRUST, which provides services for data protection compliance focusing on GDPR in the healthcare sector.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>
2. European Data Protection Board (EDPB) Guidelines 3/2018 on the territorial scope of the GDPR. Available from: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_3_2018_territorial_scope_after_public_consultation_en_1.pdf
3. Article 29 Data Protection Working Party, Opinion on anonymization techniques, Adopted on 10 April 2014. Available from: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
4. Bentzen HB, Castro R, Fears R, et al. Remove obstacles to sharing health data with researchers outside of the European Union. *Nat Med.* 2021;27(8):1329–33. doi:10.1038/s41591-021-01460-0
5. Rocher L, Hendrickx JM, de Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun.* 2019;10(1):3069. doi:10.1038/s41467-019-10933-3do
6. CNIL's referenced methodology MR-001, Research in health sector with consent. Available from: <https://www.cnil.fr/fr/declaration/mr-001-recherches-dans-le-domaine-de-la-sante-avec-recueil-du-consentement>
7. EMA Policy 0070. External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use, ver 1.4. 15 Oct 2018. Available from: <https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/clinical-data-publication/support-industry/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data>
8. The Case Reports (CARE) guidelines [cited 2022 Nov 27]. Available from: <https://www.care-statement.org>
9. Roguljić M, Ščepanović R, Rees M. Writing case reports, consent for publication and General Data Protection Regulation (GDPR). *Case Rep Womens Health.* 2020;27:e00204. doi:10.1016/j.crwh.2020.e00204

Author information

Tatiana Revenco, PhD, is a certified Data Protection Officer, with scientific background and experience in publications of scientific articles related to clinical research. https://scholar.google.fr/citations?hl=en&user=hc_8mt8AAAAJ&view_op=list_works&sortby=title



Gregory Collet, PhD, is a certified Data Protection Officer, scientific collaborator, and assistant professor in statistics at Université Libre de Bruxelles, with experience in publications and medical writing. <https://scholar.google.fr/citations?user=6iDW6VUAAAAJ&hl=en>

