

Regulatory Matters

Anonymisation and data privacy:

A medical writer's perspective on the relationship between EMA Policy 0070 and the General Data Protection Regulation



Over the last few years, the call for anonymisation has been increasing – both in frequency and volume – and like snoozing the alarm, resistance, eventually, is futile. From the perspective of a medical writer, especially one within a contract research organisation (CRO) or other types of vendors, what can we do to make sure it's performed correctly, whilst protecting data privacy and retaining the data's scientific value? In this article, we look at how the two concurrent pieces of EU legislation, EMA Policy 0070 and the General Data Protection Regulation (GDPR), need to be addressed from the perspective of the medical writer. Rather than going into detail on either, we will focus on the tensions between the two, drawing on experience from the medical writing team of a CRO.

EMA Policy 0070

Policy 0070, which was implemented in January 2015, makes it legally binding to make public the clinical information included in a marketing authorisation application. Although publicising clinical research information builds faith in the

pharmaceutical industry and aids further research, a side effect is the unwitting publication of personal or proprietary data. So, as part of the submission process under Policy 0070, any "sensitive" information needs to be protected or removed.

Policy 0070 is being rolled out in two phases. The first phase specifies that all "clinical reports" included in the submission will be made publicly available by the EMA. Clinical reports include clinical overviews (Module 2.5), clinical summaries (Module 2.7), and the clinical study reports (Module 5), plus certain clinical study report appendices: 16.1.1 (protocol and protocol amendments), 16.1.2 (sample case report form), and 16.1.9 (documentation of statistical methods). The second phase, with an unknown implementation date, will include the public disclosure of individual patient data included in the submission. In December 2018, the EMA temporarily suspended all Policy 0070 activities and the resumption date is yet to be announced. However, 131 submissions had already been authorised and are now in the public domain.¹

SECTION EDITOR



Jennifer Clemens, ELS

jennifer.clemens@merck.com

The protection or removal of sensitive information from clinical reports is achieved by anonymisation, and this is where the medical writer plays a role. The anonymisation can be done either proactively or post hoc, using redacting/masking techniques or more sophisticated, automated anonymisation techniques. More on this topic below.

Policy 0070 defines two different categories of sensitive information: commercially confidential information (CCI) and protected personal data (PPD). The CCI includes any information that isn't already publicly available and may have a financial impact on the Market Authorisation Holder if it were made publicly available. The PPD includes information relating to an identifiable person. Definitions and limits of each are detailed in Policy 0070, but the job is to identify what actually falls into the scope of each definition; this needs to be agreed upon with the key stakeholders in advance, to create a redaction strategy for the submission. It may be more likely that a CRO/vendor medical writer would be involved in helping the sponsor define this strategy than would a medical writer in the sponsor company itself.

Medical writers will need to work with biostatisticians, regulatory advisers, privacy/intellectual property associates, as well as key members of clinical and nonclinical teams, to create a strong cross-functional team. The objective is to build a predefined strategy for anonymisation of clinical data.

Policy 0070 also specifies that an Anonymisation Report is generated as part of the submission package. Among other topics, this report includes details on the anonymisation methodology and how the risk of re-identification is measured and managed. We'll look at this in more detail later, but it's quite likely that medical writers will be involved in generating this report.

The EU's General Data Protection Regulation (GDPR)

Our attention to the thorny issue of data privacy and public disclosure became more focussed after the thought-provoking presentation on Policy 0070 and the EU's GDPR, given by Elizabeth Youngkin and Raquel Billiones as part of the EMWA Expert Seminar Series, in May 2018. One of the points discussed during the presentation was the eye-watering cost of the fines that could be imposed if a data breach were to occur (4% of the annual company turnover or 20 million Euros – *whichever is greater*). The GDPR has been in force since May 2018, and a glance at Wikipedia shows us that the highest fine for any single company so far has been 183 million pounds, sterling,² just in case there was any doubt that this is serious business.

So, as medical writers, what do we need to know about anonymisation within the framework of Policy 0070 and the GDPR? A key part of the GDPR is the legal requirement to protect an individual's private data, **by design** (i.e., using built-in systems to ensure compliance) and **by default** (i.e., the minimisation of data collection, processing, and reporting). This means we need to have predefined processes and systems to ensure that:

- Only necessary data are collected and processed.
- Data are anonymised appropriately.

Also, the risk of re-identifying data needs to be assessed. As noted above, Policy 0070 provides guidance on what needs to be anonymised, as well as how to address the risk of re-identification. Privacy by design and default brings us back to the use of proactive anonymisation. Medical writing departments and medical writers will need to reconsider how to present data so that only the most clinically meaningful information is presented, whilst aiming to reduce the presence of CCI and PPD in the reports. One of the purposes of the Clarity and Openness in Reporting: E3-based (CORE) Reference is to address the need for proactive anonymisation. The objective is to think ahead to what would need to be protected to reduce the need for later redaction or masking. This can help the medical writer approach report writing with a "data protection by design and default" mindset and help comply with the GDPR. Medical writers are very familiar with the deliverables that are included under Policy 0070, are well-placed to approach the writing with "protection by design" in mind, and can perform the redaction and masking. We therefore have a very valuable role in the overall process.

Although Policy 0070 provides definitions and limits for the anonymisation, it does not mandate any particular method to achieve it. However, feedback from the first phase of Policy 0070, as well as the review process, is leading to a consensus on what needs to be anonymised. The suspension of all Policy 0070 activities offers an opportunity to fine-tune internal strategies and processes and reinforce best practices.

A risky business?

As we know, the Anonymisation Report should address the risk of re-identification. This is defined as the probability of re-identifying trial participants once they have been anonymised. Just think about how investigative journalists can uncover information about an individual by making connections through seemingly unrelated pieces of data, or how artificial intelligence is evolving and can be used to make these connections. These are so-called adversary attacks on the data – a deliberate attempt to "crack the code". Equally, re-identification can occur through unintentional discovery of an individual's identity.

Risk of re-identification can be assessed either quantitatively or qualitatively. Qualitative risk is calculated by a subjective assessment of the risk of re-identification, usually as either low, medium, or high. Quantitative risk is calculated using data to produce a numerical value, which is then used to predict the probability of re-identification. Policy 0070 recommends a threshold of 0.09 for quantitative risk of re-identification. In an informative article on the topic, Raquel Billiones³ analysed the anonymisation methodology used for the redaction packages submitted as of December 2017. Of 45 redaction packages that identified a risk of re-identification methodology, 39 used a qualitative method. So, for now, the qualitative method is the most commonly used, which fits with the use of redaction and masking. This will likely change to a predominance of quantitative methodologies in the future.

Who's responsible?

Two roles are clearly defined in the GDPR: data controller and data processor. The data controller is responsible for determining the purposes and the methodology for the processing of personal data. The data processor is responsible for processing personal data **on behalf of** the controller. What isn't clearly defined is how sponsors, vendors, and the individuals doing the

work are legally responsible in terms of breach of data. This is where it gets nerve-wracking for medical writers.

One way of looking at it that has been proposed is that medical writers are the data processors, taking clear instruction from the methodology defined by the data controller. It gets more complicated when we look at the role of the CRO vs that of the sponsor. This can be interpreted as the sponsor acting as the data controller and the CRO as the processor. The

discussion is still open, but the GDPR guidance recommends that these roles are very clearly defined up-front in the contract for the work, and the data controller is responsible for ensuring that the data processor has in place appropriate technical and organisational measures to meet the GDPR requirements.

So, coming back to we medical writers: GDPR compliance and Policy 0070 requirements are becoming intermeshed with our

approaches to data presentation. CRO medical writers should check responsibilities are clearly defined in the contract and consult with company privacy experts. Sponsor medical writers will need input from the appropriate group responsible for privacy within their company.

The take-home message? It's not solely a writer's role to decide what should and shouldn't be removed from a report – it will be a collaborative review and approval process, so make sure that you have the support you need.

It's not solely a writer's role to decide what should and shouldn't be removed from a report – it will be a collaborative review and approval process.

References

1. European Medicines Agency. Available from: <https://clinicaldata.ema.europa.eu/web/cdp/home>.
2. GDPR Fines and Notices. Available from: https://en.wikipedia.org/wiki/GDPR_fines_and_notices.
3. Billiones R. Anonymisation reports from 2016 to 2017: A preliminary analysis. *Med Writ*. 2018;27(2):22–6.

James Wolfe, PhD

Senior Director, Medical Writing
Parexel International
James.Wolfe@parexel.com

Dr Simin Takidar

Principal Medical Writer
Parexel International
Simin.Takidar@parexel.com