Pseudonymisation of Personal Data in Pharmaceutical Research: The Art of Disguising Identities

Wim Nauwelaerts explains an important concept in the field of EU data privacy, while *Linda Horton* examines issues in the US.

Respecting individuals' privacy is of cardinal importance in any type of pharmaceutical research or study that purports to consider the effects of a medicinal product on human beings, *writes Wim Nauwelaerts*. The European Union (EU) Clinical Trials Directive (Directive 2001/20/EC) therefore provides that a study subject's right to data protection and privacy must be safeguarded in the context of a clinical study. The World Medical Association's Helsinki Declaration and the International Conference on Harmonisation (ICH) Good Clinical Practice (GCP) guidelines also require those conducting clinical studies to respect the privacy of study subjects and ensure the confidentiality of personal information.

The privacy of individuals taking part in pharmaceutical studies is a recognised fundamental right and is safeguarded by EU rules that prevent personal data from being used to a study subject's detriment

Under applicable EU data privacy rules – as transposed into the laws of the different EU member states – the processing of study subjects' personal data for the purposes of conducting pharmaceutical research is subject to legal restrictions. These restrictions are aimed at fostering special protection, in particular, of study subjects' health-related information and other "sensitive" personal data, to ensure that the data are not disseminated or otherwise processed to the study subject's detriment.

Data privacy rules definitions

EU data privacy rules protect study subjects to the extent that the subjects' personal data are being processed – before, during or after the study. The EU Data Privacy Directive (95/46/EC) defines "processing" broadly as any operation or set of operations which is performed upon personal data. This includes collection, storage, disclosure etc. "Personal data" refers to any information relating to an identified or identifiable natural person. A person is "identified" when his or her identity is established, distinguishing him or her from other persons within the same group. A person is "identifiable" if he or she can be identified – directly or indirectly – by reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic, cultural or social identity.

It is therefore essential that at least one "identifier" is embedded in the data relating to a natural person in order for that data to be considered personal data under EU data privacy rules. However, in one of its recent opinions, the Article 29 Working Party – an advisory body to the European Commission comprising representatives of the national data protection authorities – emphasises that the mere hypothetical possibility of singling out an individual is not sufficient to conclude that the person is identifiable. If the possibility of identification on the basis of certain information does not exist, or the possibility is negligible taking into consideration "all the means likely reasonably to be used", the person should not be viewed as identifiable, and the information would not constitute personal data. In other words, if the person processing certain information makes use of means that are likely reasonably to result in the identification of an individual, the information itself probably qualifies as personal data.

A European Commission advisory body recently stressed that a mere hypothetical possibility of pinpointing an individual is not enough to deem the person identifiable

Pseudonymisation vs anonymisation

A popular way for recipients of data to avoid the application of EU data privacy rules is to have all personal identifiers (ie name, address, date of birth) removed from the information before it is passed on to them. Often personal identifiers are removed permanently – this is referred to as "anonymisation" of personal data. Although anonymisation offers several advantages from an EU data privacy perspective, it may not be possible or suitable in all cases. Sometimes "pseudonymisation" is preferable.

Data recipients often avoid data privacy rules using anonymisation or pseudonymisation

Pseudonymisation is the process of disguising the identities of individuals so that information relating to those individuals can be handled without knowing to whom the information relates. In common with anonymisation, pseudonymisation removes directly identifiable information. Unlike anonymisation, however, pseudonymisation is reversible. With pseudonymisation, it is possible to trace back data to the individuals by using, for instance, corresponding lists for identifiers and their pseudonyms or two-way encryption algorithms.

Unlike anonymisation, which permanently removes personal identifiers, pseudonymisation disguises identities and is reversible

Wim Nauwelaerts is an attorney at the law firm of Hogan & Hartson in Brussels and is also a member of the Brussels bar. **Linda Horton**, a former FDA director of international policy, is a partner at Hogan & Hartson, also in Brussels.

The advisory body suggests that retraceable pseudonymised information be regarded as personal data, but calls for a flexible application of privacy rules in such cases since the risk for improper disclosures are low

In line with the position taken in some EU member states, the Article 29 Working Party has suggested that retraceable pseudonymised information may be considered as personal data. Arguably the persons to whom the information relates are indirectly identifiable by someone who has the means to reverse the pseudonymisation process (even if in practice this is unlikely to occur). However, the Article 29 Working Party advocates a flexible application of data privacy rules in this case, as the actual risks of improper data disclosures for individuals (whose data have been pseudonymised) are rather low.

One example of pseudonymisation regularly applied in the context of pharmaceutical studies is the encoding or key-coding of study subjects' personal data. Personal data are earmarked by a code, while the "key" to unlock the identifiable information is kept separately, usually by the researcher or investigator conducting the study. The likelihood of this key being accessed by the sponsoring pharmaceutical company will play an important role in determining whether or not pseudonymised subject information should be regarded as personal data. The pharmaceutical company's ability to access personal data keys in particular circumstances will typically depend on the type and objective of the study.

Transfer of medical data

Observational studies

Health-related information in observational studies is often psuedonymised and the re-identification of subjects for adverse events follow-up reporting is not usually required

Some pharmaceutical research involves the transfer of medical data from hospitals or physicians to a pharmaceutical company for the purposes of conducting an observational (ie non-interventional) study. The health-related information is often pseudonymised before it is sent to the pharmaceutical company, while the names and addresses of the patients to whom the information relates stay exclusively with the hospitals or physicians. In addition, hospitals or physicians will take all necessary legal, technical or organisational measures to avoid the possibility that study subjects would be become identifiable to the pharmaceutical company.

Usually this type of study does not involve an obligation to report and follow up on adverse events or reactions, so re-identification of study subjects for that purpose would not be required. Against this background, EU data protection authorities might consider that the pharmaceutical company does not have data processing means which are likely reasonably to be used to identify the study subjects. If this reasoning is followed, the pharmaceutical company would not be receiving personal data as defined by EU data privacy law. Consequently, the company would not be considered the data "controller", ie the person determining both the purposes and the means of the data processing, and thus responsible for such processing.

Interventional studies

The situation is different in the case of an interventional study (ie a trial involving an investigational medicinal product under the Clinical Trials Directive and implementing member state laws).

Due to the health risks in interventional studies, pseudonymisation is tailored towards re-identifying subjects under certain circumstances

This type of study is essentially used to determine whether new biomedical or behavioural interventions are effective and safe for patients. Typically, the sponsoring pharmaceutical company will receive key-coded information relating to the participating study subjects, while the investigators retain the key. The investigator is obliged to keep the key in order to be able to identify study subjects if their health is in danger as a result of the study and to provide them with medical treatment. Both the ICH GCP guidelines and the Helsinki Declaration recommend that the investigator assign a unique identifier (a "subject identification code") to each study subject to protect the subject's identity and used in lieu of the subject's name if and when the investigator reports adverse events or other study-related data. In addition, the Clinical Trials Directive requires that follow-up reports for the purposes of monitoring adverse reactions shall identify subjects by their unique code numbers.

Because of the potential health risks for study subjects in this type of the study, the pseudonymisation process is geared towards the re-identification of study subjects under certain circumstances. Or, as the Article 29 Working Party put it, the identification of study subjects is embedded in the purposes and the means of the data processing. EU data protection authorities could therefore conclude that, in this case, the key-coded information constitutes personal data for all those individuals involved in the study who have the means to identify study subjects. This could include the sponsor/pharmaceutical company unless, for example, contractual or legal measures are put in place to prevent re-identification.

In sum, it is clear that pseudonymisation can be a useful tool in pharmaceutical research to protect study subjects' privacy while allowing for re-identification in exceptional circumstances. Whether pseudonymised information constitutes personal data as defined under EU data privacy law will depend on the type and objective of the study.

Depending on the study type and objective, the pseudonymisation process may serve a specific purpose aimed at re-identification when study subjects' health is at stake. In those cases,

pseudonymised information may be regarded as personal data and their processing would thus be subject to EU data privacy rules.

In a recent opinion, however, the Article 29 Working Party encourages a flexible approach based on a "likely identification" test in order to determine whether information is personal data from the perspective of a particular person who receives, handles or otherwise processes the information.

It will be interesting to see how the national data protection authorities in those EU member states that have traditionally imposed stringent rules on the processing of pseudonymised personal data will react. Will they subscribe to the Article 29 Working Party's views and apply a more flexible regime in the future? A uniform, less rigid approach across the EU would have two obvious benefits. It would enhance legal certainty for pharmaceutical companies involved in international multicentre studies and would also increase the attractiveness of the EU as a forum for pharmaceutical research.

It remains to be seen how national authorities in the EU will respond to the advisory body's call for a less rigid approach

Data privacy in the US

The above article sheds lights one of the most difficult aspects of EU law governing pharmaceutical and medical device clinical trials, ie how to meet the requirements of EU privacy law and at the same time satisfy regulatory requirements, *writes Linda Horton*.

Agencies such as the US Food and Drug Administration (FDA) may insist upon access to personally identifiable data about clinical study subjects during GCP inspections of key trial sites – sometimes in Europe – that had generated pivotal data in support of pending New Drug Applications (NDAs). The FDA rarely, if ever, conducts audits of ongoing trials but in the typical case conducts its audits to verify the adequacy and accuracy of data in pending NDAs, as part of its "trust but verify" approach to applications for marketing authorisation.

The FDA generally has no interest in information about the identity of individual patients or in receiving individually identifiable information. However, the agency believes firmly that there are times when it needs access to such information.

For example, if the FDA suspects that a clinical investigator has fabricated medical records on fictitious subjects (as has occurred when investigators found it difficult to enroll enough actual subjects), it may seek access to and copies of information needed to document this fraud. In addition, the agency sometimes wants information to verify whether a patient's profile matched the protocol's inclusion or exclusion criteria. Finally, if it appears that many subjects dropped out of the study before it was completed, and the case reports on such subjects are incomplete, the FDA's auditors might seek additional details about the reasons for subject's leaving the study because some an absence of data on subjects who dropped out for such reasons as an adverse event, or a perception that the product is not working, would undermine the applicant's case for the safety and effectiveness of the product and thus upon the agency's decision whether to approve the product.

There are situations in which the US FDA will want access to information about a subject's identity, such as in suspected cases of fraud by clinical investigators

What are the consequences of an FDA decision that it cannot trust the data coming from a particular clinical investigator? The agency might exclude the suspect data from the data set upon which the applicant hopes to gain approval. Alternatively, it might ask the sponsor to justify why the agency should consider the suspect data despite the irregularities noted by the agency's investigator during the inspection. In addition, the FDA has administrative and judicial remedies against clinical investigators who commit fraud or violate important regulatory requirements, although these remedies are rarely initiated against investigators outside the US. The FDA sometimes "disqualifies" clinical investigators who repeatedly or deliberately violate its clinical trial regulations, and the agency posts on its website lists of investigators who have been disqualified or otherwise sanctioned for such violations.

Finally, the FDA might refer a case to the US Justice Department where the agency believes that an investigator has submitted false information to a drug company sponsor knowing that the information would later be submitted to the government.