Controlling personal data in clinical trials

Trefwoorden:
clinical trial, data flows, key-coded data

This contribution follows up on an article published in Privacy & Informatie in August 2005. In the previous article, we discussed how diverging interpretations of basic data protection concepts significantly complicate data protection compliance in the clinical trials context, in particular, on a pan-European level. We indicated, for example, that the basic concept of ‘personal data’ was applied differently to key-coded data, which are generally used in the clinical trials. While in some countries key-coded data held by the sponsor were always regarded as personal data because someone could de-code the data (e.g., the investigator), in others they were not if the sponsor himself did not have the means to decode the data. We also highlighted inconsistent interpretations with respect to the responsibility of the different actors involved in the clinical trial process. Sponsors were sometimes categorized as data controllers, and sometimes not, depending on whether the data they processed were considered personal data or not. Investigators were sometimes categorized as data controllers, and sometimes data processors. We demonstrated that on a pan-European level, these divergences could lead to situations in which theoretically no data protection law would apply.

Recent developments in data protection law and its impact on clinical trials have surfaced since our first article on the topic. In the first article, we presented a description of the data flows generally occurring in the clinical trial context, which we have included again as it is important to understand those basic flows in order to best follow our analysis. This analysis updates our discussion of the definition of personal data and the effects associated with the key coding of clinical trial data – effects that will have an impact on the data protection responsibilities of the various actors involved in clinical trials. These responsibilities are discussed in the third section of this article. Finally, we will re-asses some of the conclusions we made in 2005.

1 Data flows in clinical trials

The graphic below describes the data flows in clinical trials, as produced in our previous article.

Physicians (also called ‘investigators’) collect the personal data of patients in the course of providing regular medical care. For patients who consented to participate in a clinical trial, the investigator incorporates (some of) this data and additional trial-specific data in Case Report Forms (CRF’s). The identity of patients is usually not disclosed in the CRF’s; their names are replaced by an alphanumerical code, for example, on the basis of their initials and a random number or birth date.

The CRF’s are shared with the ‘sponsor’ of the study, often a pharmaceutical company, for further review and statistical analysis. Because of the need to maintain data integrity, the conduct of the trial and the information included on CRF’s is subject to rigorous control.

Sponsors often use ‘clinical research organizations’ (CRO’s) to monitor the trial. CRO’s have access to the research sites and source documents (such as patient medical records), and can verify the accuracy of the data included in the CRF’s and other aspects of the trial implementation. Both sponsors and CRO’s may turn to service providers for certain aspects of the trial.

The relationships between patients, investigators, sponsors, and CRO’s are governed by a variety of instruments, such as patient informed consents, research protocols, and contractual arrangements. In addition, the conduct of a trial is subject to supervision by ethics committees and public medical inspection bodies.

2 Definition on personal data and the status of key-coded data

The definition of personal data has received a tremendous amount of attention over the last couple of years, not least because the issue has raised much controversy in the area of information technology applications and practices, such as RFID, internet anti-piracy efforts, and online customer profiling.

In clinical trials, health data incorporated in CRF’s are
Our 2005 contribution demonstrated that data protection laws and regulators adopt diverging positions on the status of such key-coded trial data, and indirectly identifiable data in general. At least one fundamental point of contention relates to the data protection directive’s provision that, in assessing the status of this data, one should consider the means reasonably and practically available to identify data subjects, taking into account the cost and time involved in that effort. However, regulatory authorities seem to disagree on who those means of identification should be available to: whether solely the entity holding the data or anyone else (as provided in recital 26 of the directive). In other words, to understand if the key-coded data held by the sponsor are to be treated as personal data, should we consider only the sponsor’s ability to identify the patient (using reasonable means), or that of the investigator as well? While this might seem very technical, the point is fundamental for qualifying key-coded data either as personal data subject to data protection law, or anonymous data, which falls outside the scope of relevant regulation.

If only the sponsor’s ability to identify a patient is relevant to whether key-coded data qualify as personal data or not, one could conclude that the sponsor is unable to identify the patient if the data are adequately key-coded and no other factors allow for easy identification (e.g., patients with particular characteristics). If the ability of the sponsor and anyone else is relevant, key-coded data in the hands of the sponsor will always qualify as personal data because someone (i.e., the investigator) is able to de-code the data, and quite easily.

Below we set out the most important recent developments in this debate, such as the recent Article 29 Working Party opinion, and guidelines adopted by the Italian and Portuguese data protection authorities.

2.1 Article 29 Working Party

The Article 29 Working Party, which is a group composed of each of the EU Member States’ independent data protection authorities, adopted an opinion on the definition of personal data in June 2007. The opinion discusses the various components of the term ‘personal data’ including its reference to ‘identified or identifiable’ natural persons – the most important element of our discussion.

The Working Party discusses at length the complex issues surrounding data on indirectly ‘identifiable’ natural persons, and attempts to set out some basic rules as to when such data are to be considered personal data and thus subject to data protection law. In particular, the Working Party discusses recital 26 of the directive, which provides that, in order to evaluate the nature of indirectly identifiable data, one should consider ‘all means likely reasonably to be used either by the controller or anyone else.’ This means, according to the Working Party, that all factors at stake should be taken into account, including the intended purpose behind the collection of the data, the structuring of processing operations, the benefits for the controller, the effects on the data subjects, and available state-of-the-art technological processes to identify individuals on the basis of available data.

The importance of the ‘purpose’ pursued by the entity that holds the data is particularly stressed. Indeed, if data are being held for purposes of identification, it is difficult to argue that the data are not personal data. According to the Working Party, technical measures preventing the identification of data subjects, such as key-coding, can thus serve two distinct objectives depending on the purpose for which the data are used:

1. Either the purpose includes the possible identification of the data subject, in which case the technical measure (key-coding) is a security measure – a ‘consequence’ of article 17 of the directive – and the relevant data remain personal data; or
2. The purpose does not include the possible identification of the data subject, in which case the technical measure is a ‘condition’ for the data not to be considered personal data.

Two practical examples demonstrate this analysis. First, a company receiving key-coded patient data from a physician (presumably in a registry) may be considered not to be processing personal data provided the data alone do not allow for patient identification, and adequate measures (technical, legal and organizational) have been taken to prevent the company from identifying the patients. While the example does not indicate explicitly, it seems to assume that the company will never have any reason (purpose) or possibility to have the code reversed and identify the patient.

The second example discusses the use of key-coded data in clinical trials. In contrast to the previous example, the Working Party now holds that the key-coded trial data should generally be considered personal data. The difference is that key-coding in clinical trials is set up precisely with the objective of enabling the identification of a patient when necessary:

The pharmaceutical company has construed the means for the processing, including the organizational measures and its relations with the researcher who holds the key in such a way that the identification of individuals is not only something that may happen, but rather as something that must happen under certain circumstances.

However, the Working Party points out that this key-coded trial data should not always be considered personal data. Indeed, the same key-coded data may be used by entities that will never have the need, interest, or ability (because of

3 The explicit conditional tense (‘may’) in the example seems to indicate that not all data protection authorities necessarily share this assessment.
CONTROLLING PERSONAL DATA IN CLINICAL TRIALS

While the Working Party does not offer examples, we believe this could cover cases where a sponsor shares the key-coded trial data with a partnering company cooperating in the research on the development of a drug. If this partnering company has no formal role in the trial (and thus no responsibility), it will have no reason to ever identify the patients, and may thus be considered not to be processing personal data.5

In the case of clinical trials, the Working Party’s bottom line seems to be that trial data should be considered personal data for those who may have access to the key (or non-key-coded source data) in certain circumstances, thus to most parties directly involved in a trial, such as the sponsor, the CRO, and obviously the investigator.

2.2 France

In January 2006, the French data protection authority (CNIL) adopted a ‘reference procedure’ for clinical trials.6 The document facilitates the CNIL authorization process for clinical trials as required under the French data protection law.7

Sponsors who formally commit to comply with the conditions set out in the reference procedure should make a simple notification to that effect, upon which they receive an automatic authorization.

The reference procedure sets out detailed rules on how clinical trial data can be collected and used by sponsors, including rules on information given to patients, disclosures of data, rights of patients, etc. The procedure provides that personal data can be key-coded by using alphanumerical numbers (consisting of the three first letters of the patient name, or, better still, the patient’s initials) and that sponsors may only receive such key-coded data and no nominative data. The rationale of the procedure logically infers that the data protection law continues to apply to key-coded data collected by the sponsor.

Finally, it is worth noting that CNIL officials have indicated that the proposed key-coding mechanism in the procedure is probably a bit weak and that the CNIL will endeavor to convince companies to apply more secure key-coding measures. However, there is no indication that those measures would affect the legal status of the data. Rephrasing the Article 29 Working Party’s position on the objective of key-coding: key-coding under the French clinical trial procedure is thus mainly a security measure, a consequence of article 17 of the directive, and not a condition for the data not to be considered personal data. While this is consistent with the Article 29 Working Party opinion discussed above, the French interpretation of the concept of personal data does not always seem to be in line with it.

As discussed in our previous article, the CNIL, and its Belgian and Swedish counterparts, adopt a very expansive definition of the term ‘personal data’ that seems to rely heavily on the possibility of anyone identifying the data subject, regardless of the purpose pursued, or the safeguards in place. As a result, any data coded by a key-coding mechanism that can be reversed by anyone, using reasonable means, remain personal data. Under this interpretation, key-coding is always a consequence of article 17 of the directive, and never a condition for the related data not to be considered personal data.8

2.3 Italy

In July 2008, the Italian data protection authority (Garante) adopted guidelines on the processing of personal data in clinical trials.9 The guidelines consider that key-coded patient data held by a sponsor are personal data. According to the Garante, various arguments support this assumption, including the following: (i) the sponsor, through its service provider, the CRO, may have access to non-coded source data; (ii) the sponsor is likely to receive sufficient information to be able to identify a patient regardless of the key-coding; (iii) the key-coding is performed with the explicit intention of enabling the identification of a patient in particular circumstances (for example to interrupt a therapy); and (iv) the investigator has a list that makes the data identifiable, regardless of whether the sponsor has access to it.

The Garante thus seems to combine the two interpretations set out above: first, the sponsor itself is likely to have sufficient data itself to identify the patient; and second, even if the sponsor does not, the data are still personal data because the investigator can identify the patient. The Garante explicitly qualifies key-coding as a security measure, which in and of itself, does not affect the legal status of the data. As a result, the Garante concludes that key-coded data held by the sponsor should be considered personal data, and, more precisely, sensitive personal data.

2.4 Spain

In September 2008, the Spanish data protection authority adopted an opinion on the processing of personal data in clinical trials that assesses the qualification of key-coded trial data under the data protection law.10 The authority concludes that the law applies to key-coded clinical trial data, ‘even if some entities involved in the trial only have access, in principle, to key-coded data.’ While the opinion itself does not specify it, the authority is presumably referring to trial sponsors.

6 CNIL, Méthodologie de référence pour les traitements de données personnelles opérés dans le cadre des recherches biomédicales, January 2006.
7 Loi du 6 janvier 1978 relative à l’informatique, aux fichiers et aux libertés, article 54.
8 This could explain the conditional tense referred to in footnote 3. In the French interpretation, the data submitted by physicians in that example, would still be personal data, unlike what the Working Party suggests.
2.5 Portugal
In July 2007, the Portuguese data protection authority adopted a decision on the processing of personal data in clinical trials.11 The decision does not discuss the status of key-coded data in particular but seems to consider it self-evident that a sponsor processes personal data. Key-coding of data is discussed as a security measure to be applied to personal data but not as a measure that could change the status of data under data protection law.

2.6 Germany
The German data protection law contains a specific provision on ‘pseudonymising’ personal data (pseudonymisieren). Pseudonymising is defined as replacing a name or other identifying elements by a code with the purpose of rendering the re-identification of the individual impossible or very difficult.12 The law’s data minimization principle mandates the use of anonymous or pseudonymous data whenever possible,13 but it does not prohibit the re-identification of the individual based on the code, and does not clarify how the process of pseudonymising may affect the nature of the data under data protection law. Does the pseudonymized data remain personal data or not, and when or when not?

While authors seem to disagree on this point,14 in practice, regulatory authorities generally do not take positions in principle and prefer assessing the status of pseudonymous data on a case-by-case basis. In doing so, they consider whether the holder of key-coded data can identify a patient using means reasonably and practically available.15 If the holder of the key-coded data is unlikely to have access to the key, and cannot identify patients on the basis of the key-coded data itself, the holder may not be subject to the data protection law. Thus, in contrast to the French law, key-coded data could be considered not to be personal data, even if a key by which the coding can be reversed exists. Rephrasing the Article 29 Working Party: reversible key-coding under German law is thus not just a security measure, a consequence of article 17 of the directive, but can also be condition for the data not to be considered personal data. In the clinical trials context, however, the pseudonymized data will often be considered personal data because of the sponsor’s ability to access medical records and the potential circumstances in which the code must be broken to identify a patient.

In light of the data minimization principle, entities should use pseudonymized data where they can. In the case of clinical trials, this obligation is reinforced by the Medicines Law, which mandates the use of pseudonymous data for research purposes and for the reporting of adverse events in clinical trials.16 These obligations forced regulators to consider conditions under which data can be considered properly pseudonymized. The data protection law provides that re-identification should be impossible or very difficult. But what does this mean in practice?

Over the last couple of years, German regulatory authorities have increasingly focused on the mechanism used to key-code health data. For example, regional authorities in Bavaria,17 Brandenburg,18 and Mecklenburg-Vorpommern19 have expressed reservations about the use of initials and date of birth as coding mechanisms because this method would allow for a fairly easy identification of patients – in which case the key-coded data would not qualify as pseudonymous data and the data minimization principle may be violated. The issue has also been debated in the Working Group on Science and Research (Arbeitskreis Wissenschaft und Forschung), which regularly brings together German regional data protection authorities in an attempt to streamline their positions. While the working group has not adopted a formal position, a majority of members apparently considers that patient initials or full birth date, in principle, should not be used.20

2.7 The Netherlands
The Netherlands is one of the countries that only appears to consider the ability of the sponsor to identify a patient as relevant (not the ability of any other person). Indeed, the NEFARMA code of conduct, approved by the data protection authority, contained a sweeping statement saying that key-coded data are not personal data for the sponsor (assuming that only the investigator holds the key). The code expired in September 2007, and, according to our information, NEFARMA is still negotiating with the DPA on an extension of the code. However, the DPA apparently has asked for more detailed commitments in terms of the coding of the data. According to the DPA, the standard practice of using initials for patients in research systems is not acceptable under Dutch law and the code needs to be updated to reflect the status of pseudonymized data for the entities who do not have access to the key. According to our information, the code is expected to be renewed in the near future.

11 Comissão Nacional de Protecção de Dados, Deliberação no. 333/07 Sobre a protecção de dados pessoais nos ensaios clínicos com medicamentos de uso humano.
12 §3a BDSG.
13 §2a BDSG.
14 Johann Bizer suggests that pseudonymous data should be considered personal data. He refers specifically to the research context, where key-coding is used precisely to enable the identification of the individual when necessary (cfr. Article 29 Working Party position). According to Bizer, pseudonymous data should be considered personal data as long as the key exists. He thereby rejects the position held by Alexander Roßnagel and Philip Scholz who claim that pseudonymous data should not be considered personal data for the entities who do not have access to the key. (J. Bizer in: S. Simitis, Bundesdatenschutzgesetz, Nomos Kommentar, 6th edition, p. 315, and A. Roßnagel and P. Scholz, Datenschutz durch anonymität und pseudonymität, Multimedia und Recht 2000, p. 726.)
16 § 40(2)-a Arzneimittelgesetz (AMG).
20 Apparently, however, the working group tentatively accepts the use of full birth date in specific circumstances because this is the only way in which trial data (collected under §40(2)-a AMG) can be linked to data collected at a later stage in the context of pharmacovigilance reports (§63b AMG), for which the working group applies. It only does not accept the use of the full birth date in order to avoid double reports.
and birth dates would not suffice to make patient data anonymous, and therefore, exclude the applicability of the data protection law. To our knowledge, however, the DPA is still prepared to accept that adequately key-coded data in the hands of the sponsor could, in specific circumstances, fall outside the scope of the law, even if the investigator can reverse the code.

3 Who controls whom and what

Once the applicability of the data protection law has been established, the question of responsibility must be considered. While patients are obviously data subjects, the role of the sponsor, investigator, and CRO is more difficult to determine.

3.1 Sponsor

The responsibility of the sponsor will, in part, be determined by the data it collects and uses. As indicated above, key-coded data increasingly appear to be considered as personal data, either because of the amount of data available to the sponsor, the characteristics of the (regulatory) environment in which trials are conducted, or the investigator’s ability to reverse the code. As a logical result, sponsors will increasingly be considered data controllers subject to the data protection law. This is clearly the position taken in the French, Italian, and Portuguese documents referenced above.

In addition to the status of the key-coded data, the Italian and Portuguese positions also refer to the sponsors’ involvement in the set-up of trials, their control over the use of service providers, such as CRO’s, and their general ability to determine the means and purposes of the processing operations associated with clinical trials.

3.2 CRO

CRO’s are generally hired by the sponsor to perform a variety of functions. In some cases, their role is limited to a monitoring function, in which they mainly verify the proper implementation of the trial by the investigators and their teams, and the accuracy of the data included in ICF’s. In other cases, they play a more prominent role and are paid by the sponsor to conduct most aspects of the trials. As indicated in our 2005 article, CRO’s are frequently considered data processors of the sponsor. The new French, Italian and Portuguese texts support this position.

3.3 Investigator

As indicated in our previous contribution, the position of the investigator is a delicate one. On the one hand, the investigator is subject to very strict rules, controls, and contractual obligations imposed by the sponsor, which all seem to support a ‘data processor’ qualification. On the other hand, investigators have their own professional responsibility, have a relationship with the data subject, and are actively involved in the set-up of a trial, all of which support a controller qualification. In our experience, the qualification of the investigator as data (co-)controller is most common.21 This has recently been supported by the Garante in its consultation document stating the individual trial facilities and sponsor companies in general have distinct responsibilities within the field of clinical studies and therefore amount to autonomous actors.’ This conclusion, however, does not seem to be supported by the 2007 Decision of the Portuguese DPA. Indeed, the Portuguese DPA suggests that the investigator is a data processor of the sponsor.22

4 Conclusion

The developments discussed above demonstrate an increasing convergence in the positions taken by regulators. In summary, these positions hold that (1) key-coded clinical trial data processed by sponsors qualify as personal data; (2) the investigator and sponsor are both considered data controllers in their own right; and (3) generally CRO’s are data processors of the sponsors.

In our experience, this general fact pattern roughly reflects how most established pharmaceutical companies have approached their clinical trials to date. Yet, this by no means resolves all the issues. For example, the qualification of key-coded data as personal data, and the full application of the data protection law to that data, raises significant concerns. The Garante’s position is illustrative: while the Garante apparently feels the need to explain at length why key-coded trial data qualify as personal data, it does not hesitate to then immediately flag the data as sensitive, and thus, subject it to the most stringent regulatory regime. This would seem somewhat disproportionate. If the legal status of the data is unclear, one would not expect the data to then be automatically categorized as sensitive. The Garante seems to adopt a very legalistic approach, not taking account of the limited risks associated with processing key-coded data and the highly regulated environment in which the data are used. Areas in which such a strict approach is bound to raise problems include, for example, the further use of key-coded data for unanticipated scientific research, and the further use of coded body samples and data derived from them.

In our previous contribution, we stressed the need to strike a balance between the risks involved in the processing of key-coded health data and the obligations imposed on pharmaceutical companies. We also suggested that an EU code of conduct may be a useful tool to that effect.

Up to now, however, the concept of a code of conduct has not gathered any meaningful support. Two mutually enforcing reasons come to mind. First, the pharmaceutical industry seems to believe that industry practices are too divergent to come to a reasonably supported industry position. Second, the legal environment is often considered to be too diverse for

21 Note that in countries where sponsors would not be considered data controllers because the key-coded data they collect are not personal data, the investigator could not be a data processor (because there is no data controller), and would automatically become a data controller.

22 Comissão Nacional de Protecção de Dados, Deliberação no. 333/07 Sobre a protecção de dados pessoais nos ensaios clínicos com medicamentos de uso humano, paragraph IV(5).
all data protection authorities to agree on a single code. There is certainly some truth to both arguments, and a broad code covering all aspects of the pharmaceutical sector’s activities is indeed unlikely to be an achievable objective in the short term.

However, as regulators’ positions and understandings of key aspects of the pharmaceutical sector’s activities continue to converge, a code may become easier to achieve. In practice, we believe there is room for a code, be it in discrete, well-defined areas (e.g., pharmacovigilance). More so, we believe the time is right. As regulators are increasingly focusing on the sector and taking positions on how the sector should approach compliance, it will be important for industry to ensure, first, that this increased convergence is sufficient to allow for a more harmonized compliance regime throughout the EU, and, second, goes in the right direction and does not inhibit certain uses of data essential for the research and development of new medicines and devices.

23 See, for example, the CNIL’s procedure for pharmacovigilance, adopted in January 2008.