

**INTERNATIONAL PHARMACEUTICAL PRIVACY CONSORTIUM COMMENTS
ON ART. 29 DATA PROTECTION WORKING PARTY OPINION 4/2007
(ON THE CONCEPT OF PERSONAL DATA) AND RELATED ISSUES**

EXECUTIVE SUMMARY

- The International Pharmaceutical Privacy Consortium (IPPC) commends the Article 29 Data Protection Working Party for recognising the need to promote a clearer understanding of personal data (*Opinion 4/2007 on the Concept of Personal Data*) and appreciates this opportunity to comment on the Opinion. IPPC members are research-based pharmaceutical companies that can be a resource to the Working Party by explaining the application and impact of data protection regulations on medical research and public safety reporting activities.
- Maintaining data confidentiality and subject privacy are essential to clinical research and pharmacovigilance. Privacy-protecting safeguards have been incorporated throughout the extensive regulatory systems under which pharmaceutical companies must operate and have been integrated into pharmaceutical processes. Appropriate implementation of the Data Protection Directive serves to complement existing protections.
- Pharmaceutical companies are responsible for the safety of their products, and have ethical, legal, and regulatory obligations to accurately collect, analyze, and report adverse events in a timely fashion both during clinical trials and after a drug is on the market. Internationally recognized pharmacovigilance practices are used to meet these obligations and comply with the requirements and expectations of health authorities. Controls on the collection, use, and transfer of personal data for these purposes should not be disproportionate to the substantial public interest in conducting effective pharmacovigilance in order to protect public health and safety.
- Companies also adhere to international standards for acceptable clinical research. Technical and organisational controls have been implemented to limit access to identified and/or identifiable patient information to those members of a pharmaceutical organisation who require it. The scientific value of clinical data to internal company pharmaceutical researchers thereby can be maintained without compromising patient privacy. It is critical to differentiate those roles within a pharmaceutical organization that require access to identified and/or identifiable data in order to meet regulatory requirements from those that do not. In fact, it is rarely the case that pharmaceutical company researchers have access to directly identifiable data about patients. The requirements for processing of personal data should be applied only to those parts of a sponsor's organization that have access to identified or identifiable data.

- Data derived from biological samples can be identifiable or non-identifiable, just as other personal health data can be. Genetic data are non-identifiable unless a reference database or similar available record source exists that links the data to individual identities and to which the researcher has access. The IPPC recommends that in a future document, the Working Party, in consultation with other European institutions, distinguish between data derived from biological samples that relates to an identified or identifiable person and data that does not.
- In determining whether data are, in fact, “identifiable”, the Directive states that account must be taken of all the means reasonably likely to be used by the controller or by any other person to identify the individual.¹ The IPPC agrees with the Working Party that whether data is indirectly identifiable may depend upon the adequacy of technical and organisational controls to protect against identification. For example, research data containing dates and geographic codes should not be considered identifiable provided appropriate controls protect against patient identification.
- Data protection authorities and patients alike must have confidence that medical research and pharmacovigilance data are collected and processed in a way that respects the rights of data subjects. It is equally important to understand the way data is used during medical research and pharmacovigilance activities and to appreciate the real-world implications of greatly restricting or prohibiting certain data flows that are vital to live-saving research and safety reporting activities. In these areas critical to public health, the IPPC urges the Working Party to initiate a dialogue with national health authorities to ensure that the data protection and public health functions of government are not at odds. The IPPC looks forward to continued communications with the Working Party and other stakeholders on this subject.

¹ Directive 95/46/EC at preamble, ¶ 26.

INTRODUCTION

The International Pharmaceutical Privacy Consortium (IPPC) is an organization formed in 2002 and comprised of chief privacy officers and other privacy professionals from 15 research-based global pharmaceutical companies, all of which conduct business in the European Union and some of which also are headquartered here. Membership and mission of the IPPC is described in Attachment A. These comments are submitted on behalf of the following IPPC companies: Abbott Laboratories, AstraZeneca Pharmaceuticals, Bristol-Myers Squibb, Elan Pharmaceuticals, Inc., Eli Lilly and Company, GlaxoSmithKline, Johnson & Johnson, Merck & Co., Inc. (operating as Merck Sharp & Dohme in most countries outside USA), Pfizer Inc., Roche, sanofi-aventis, Schering-Plough, Takeda Pharmaceuticals, and Wyeth.

The IPPC sees great value in the Article 29 Working Party's efforts to analyze and further elucidate the concept of "personal data". National data protection authorities, multinational entities, and data subjects all will be well-served by increased clarity and uniform pragmatic interpretations of this central concept under the Directive and national data protection laws. The IPPC acknowledges that interpreting "personal data" in the context of medical research is a difficult task. We seek to foster increased understanding of medical research and related activities of the pharmaceutical industry and to work together with data protection scholars and regulators to ensure that privacy is appropriately protected without unduly restricting activities that benefit public health and medical advances.

The Working Party's Opinion has profound implications for the pharmaceutical sector. By submitting these comments, the IPPC hopes to clarify certain factual references in the Opinion about some types of pharmaceutical activities and the associated use of data. We also understand that the Working Party intends to address the issue of clinical research shortly, and therefore we seek to provide input on the interpretation and protection of "personal data" in the pharmaceutical research and regulatory context. Included in this is the issue of pharmacovigilance, which we introduce in these comments. The IPPC would be pleased to further discuss these and other relevant issues with the Working Party.

PHARMACOVIGILANCE

Pharmacovigilance is the science of activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem (WHO). Adverse events include a range of negative or unexpected reactions to a drug – from relatively minor irritations to potentially life threatening conditions. As required by regulations in the EU and elsewhere, in order to safeguard patient safety, pharmaceutical companies must apply internationally recognized Good Pharmacovigilance Practices (GPvP) during drug development and after obtaining marketing authorization. Good Pharmacovigilance Practices are followed in the collection, analysis, and communication of safety information to patients, healthcare practitioners, consumers, and regulators.

Access to identifiable information may be required for GPvP purposes. Examples of uses of identifiable information during GPvP include:

1. To enable contact with patients or adverse event reporters (including healthcare professionals) to ensure appropriate treatment is given as promptly as possible. The pharmaceutical company may have access to information that may not be known to the patient/reporter/treating physician. It is critical that this information can be communicated quickly to help remediate any adverse events.
2. To obtain additional information necessary for analysis of possible safety issues. For example, further information might be necessary in order to determine the clinical/biological pattern of the adverse event and identify circumstances that could increase the risk of its occurrence. Such additional information is typically obtained through active contact (follow-up) with patients, healthcare professionals, or others. These follow-up attempts are required by regulation and are standard components of GPvP.
3. To meet the regulatory requirements of various health authorities around the world who require specific information in order to consider an adverse event report to be valid. ICH Guideline E2D requires one or more of the following: age (or age category, e.g., adolescent, adult, elderly), gender, initials, date of birth, name, or patient identification number.
4. To compare newly received adverse event reports with previously received reports, for the purposes of identifying duplicate cases. Identifying duplicate cases is important to avoid overestimating the incidence of specific events either by company or regulatory safety experts.

A number of technical and organisational controls typically protect pharmacovigilance data from unauthorized access, use, alteration, loss, disclosure or other processing. It is standard practice for pharmaceutical companies to have separate groups within their organisation that are responsible for pharmacovigilance as well as separate files and databases to support these activities. The employees of the company who are responsible for pharmacovigilance activities are bound by obligations of confidentiality covered by the company's employment contracts, policies or standard operating procedures. Even within a pharmacovigilance group, confidential information learned in the course of such activities is shared only as necessary to conduct activities such as statistical analyses and regulatory reporting. In all cases, these activities are subject to rigorous health regulatory controls apart from the Data Protection Directive. These regulatory controls require that (i) access to systems containing pharmacovigilance data be restricted to those who require it in order to perform job functions; (ii) audit trails be maintained that track all database changes; and (iii) systems undergo validation to ensure accuracy, reliability, and consistent intended performance. These controls are subject to inspection by health regulators.

Inconsistency between member states (and sometimes even between health and data protection authorities within a single member state) as to whether certain data elements are "identifiable" under the Data Protection Directive has created significant compliance obstacles for pharmaceutical companies. For example, some member state health authorities require that initials and dates of birth be reported by companies when transmitting case reports.

Nevertheless, some data protection authorities have indicated to companies that in order to comply with the Directive, initials and/or dates of birth should not be transmitted absent the subject's explicit consent. It is often not possible to obtain a data subject's consent to the transfer of such information within the timeframe mandated for filing adverse event reports.

Volume 9A of the Rules Governing Medicinal Products in the European Union was intended to harmonise pharmacovigilance reporting requirements across the EU. Volume 9A provides a set of recommendations concerning the electronic transmission to the EEA's EudraVigilance database of adverse event case reports.² The recommendations state that in circumstances in which initials are known to the sender but cannot be transmitted due to data privacy requirements, the field for "initials" should be populated with "PRIVACY". The recommendations also state that age can be reported rather than a date of birth. Nevertheless, in addition to reporting to EudraVigilance, companies are also required under national legislation to report cases directly to member state competent health authorities. In some member states, the reporting of initials and/or dates of birth is required.

To resolve this dilemma, the IPPC suggests the following:

- Pharmacovigilance data that contains only dates of birth and/or patient initials should not be considered "personal data" subject to the requirements of the Directive where companies have implemented appropriate technical and organisational controls to protect the data from unauthorized access, use, alteration, loss, disclosure or other processing and the data is being used for a legally required, regulatory purpose. Such controls prevent the data from being used to identify the subject of the information.
- In the alternative, the Article 29 Working Party should confer with member state health authorities to develop an approach that is consistent across the EU. For example, health and data protection authorities might agree that initials and/or dates of birth should be collected by companies in order to identify duplicate cases but that these data elements need not be transmitted to national health authorities.
- Information on concurrent conditions, even rare conditions, should not be considered identifiable, provided companies have implemented appropriate technical and organisational controls to prevent subject identification.

DRUG RESEARCH AND DEVELOPMENT

Clinical Trials

Clinical trials are studies designed to evaluate the safety and efficacy of investigational medications by monitoring their effects on healthy volunteers or patients. Clinical trials are often randomized and controlled, meaning that the effects of an investigational medication are

² See http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-9/pdf/vol9_2007-07_upd07.pdf, page 164.

studied by comparing a group of patients receiving the treatment to a group not receiving the treatment (e.g., a placebo). The assignment of patients to each treatment group is random. This type of research necessarily involves direct interaction between the clinical trial investigator and research subjects. It also involves both the use of existing health-related data of research subjects (generally, to determine if the individual meets the inclusion/exclusion criteria for the study) and the creation and examination of new data. The IPPC has prepared the attached Clinical Research White Paper that provides detailed background on how clinical trials are conducted and the privacy controls that are incorporated into study procedures (see Attachment B).

The protection of patient privacy and data confidentiality in the conduct of clinical trials has long been a cornerstone of human subjects research. Pharmaceutical companies adhere to strict protections that have been incorporated into international standards for acceptable clinical research. For example, Good Clinical Practice guidelines adopted by the International Conference on Harmonisation³ specify that “[t]he confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with the applicable regulatory requirement(s).”⁴ Moreover, “[f]reely given informed consent should be obtained from every subject prior to clinical trial participation,” and “[a] trial should be conducted in compliance with the protocol that has received prior institutional review board (IRB)/independent ethics committee (IEC) approval/favourable opinion.”⁵ Above all else, “foreseeable risks and inconveniences should be weighed against the anticipated benefit for the individual trial subject and society. A trial should be initiated and continued only if the anticipated benefits justify the risks.”⁶

These protections have been incorporated into Directive 2001/20/EC of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use (“Clinical Trials Directive”). Indeed, the Clinical Trials Directive states that “Good clinical practice is a set of internationally recognized ethical and scientific quality requirements which must be observed for designing, conducting, recording and reporting clinical trials that involve the participation of human subjects.”⁷ It also states that “The person participating in a trial must consent to the scrutiny of personal information during inspection by competent authorities and properly authorised persons, provided that such personal information is treated as strictly confidential and is not made publicly available,”⁸ and it makes specific reference to the Data Protection Directive as a source of law that must be complied with in all circumstances.⁹

³ Available at <http://www.emea.europa.eu/pdfs/human/ich/013595en.pdf>.

⁴ Id. at 2.11.

⁵ Id. at 2.9 and 2.6.

⁶ Id. at 2.2.

⁷ Directive 2001/20/EC at Art. 1, ¶ 2.

⁸ Id. at preamble, ¶ 16.

⁹ Id. at Art. 3, ¶ 2(c).

Because the Data Protection Directive allows the processing of personal data pursuant to the explicit consent of the data subject¹⁰, existing good clinical practice standards and the requirements of the Data Protection Directive have elaborated and strengthened each other. A detailed explanation of the types of data that will be collected and transferred, the uses of those data, and the persons with whom the data will be shared and have access (including individuals acting on behalf of the sponsor) has been incorporated into the rigorous clinical study informed consent process. Each Independent Ethics Committee overseeing a study is tasked with ensuring that researchers and the study sponsor have adequate mechanisms in place to protect subject privacy and data confidentiality.

Personal data is further protected by restricting the inclusion of identifiers in reporting study results to sponsors. ICH GCP requires that unambiguous subject identification codes must be used in reporting data for each subject.¹¹ A subject identification code is “a unique identifier assigned by the investigator to each trial subject to protect the subject’s identity and used in lieu of the subject’s name when the investigator reports adverse events and/or trial-related data.”¹² Therefore, case report forms (CRFs), the mechanism by which investigators report results to pharmaceutical sponsors, contain subject identification codes instead of directly identifiable patient information such as name and address. The subject identification code list must be maintained “in a confidential manner” in the files of the investigator/institution.¹³ The investigator also maintains on-site all source documents and records, such as pertinent hospital records, clinical and office charts, laboratory notes, etc.

Pharmaceutical company researchers who review and analyse CRFs generated in clinical trials do not have access to directly identifiable data about patients. This contrasts with external clinical investigators who actually undertake clinical trials sponsored by pharmaceutical companies. Investigators are neither employees nor agents of the sponsor. They are health care professionals independent of the sponsor who have contractually agreed to conduct a clinical trial at a trial site. Although often affiliated with academic medical centres, they may be solo practitioners or providers in an organised health care arrangement. Recognition of these distinct roles of pharmaceutical company researchers and external clinical investigators will facilitate the appropriate interpretation of the Directive in this complicated area.

The external clinical investigator is, of course, at all times aware of the patient’s identity. However, the investigator is bound by a health care provider’s duty of medical confidentiality to the patient, as well as the confidentiality provisions of the study protocol, not to reveal this information. What the investigator may not know, however, is whether a particular patient is in the experimental or control arm of the study. Namely, an investigator in a double-blinded study does not know whether a particular trial subject is receiving the investigational product, a comparator product, or a placebo. “Unblinding” may be necessary in the event of a medical emergency for a trial subject. Breaking the blind involves procedures specified in the study

¹⁰ Directive 95/46/EC at Art. 8, ¶ 2(a).

¹¹ ICH Guidelines, supra note 3 at 5.5.5.

¹² Id. at 1.58.

¹³ Id. at 8.4.3.

protocol that allow the investigator to find out which treatment or placebo the patient has received.

The Opinion states that key-coded data disclosed to a pharmaceutical company sponsor by a clinical researcher who holds the key should be considered personal data where re-identification is “embedded in the purposes and means of the processing.” (p. 20). In most clinical trials, although re-identification *by the clinical investigator* is possible for medical, safety and regulatory purposes, re-identification by the pharmaceutical company sponsor is excluded in the design of protocols and procedures. Pharmaceutical sponsors are by design unaware of the patient’s identity, with two limited exceptions discussed below.

First, specific individuals employed by, or acting on behalf of, a pharmaceutical sponsor may have access to identified and/or identifiable information in the context of pharmacovigilance reporting and investigation. This issue has been discussed above; however, it is important to reiterate here that these activities are undertaken by individuals whose responsibilities are centred on pharmacovigilance requirements. Technical and organisational controls are used to protect the confidentiality of the information and to limit its broader dissemination within the organisation.

Second, a field monitor employed by the sponsor may be directed to review the source data at the study site. In accordance with ICH GCP, it is the monitor’s responsibility to “check[] the accuracy and completeness of the CRF entries, source data/documents, and other trial-related records against each other.”¹⁴ In the course of this review with the clinical investigator, field monitors may have access to identified patient information at the study site. Monitors do not remove this information from the study site and are not permitted to share this information within the sponsor organisation. Accordingly, such identified patient information that an individual monitor may see at the study site is accessible only for a limited duration for a specific purpose, and it is not transferred or included in any sponsor files or databases. Standard procedures prohibit broader dissemination because, among other things, such a disclosure could compromise good clinical practice.

The IPPC proposes that by appropriately segregating access to identified or identifiable data within the sponsor’s organisation through the implementation of adequate security measures and audit controls, the requirements for processing of personal data should be applied only to those parts of a sponsor’s organisation that have access to identified or identifiable data. This issue takes on increased importance in the context of “secondary research”, discussed below, for which a company has limited ability to obtain a data subject’s specific consent.

“Secondary Research”

As the term suggests, secondary research involves the analysis of key-coded data collected in prior clinical or other research studies for additional research purposes. These additional purposes could, for example, involve further examination of the disease or condition

¹⁴ ICH GCP, *supra* note 2, at 5.18.4(m).

in question, or examination of some unanticipated, secondary benefit of an investigational drug. Because secondary research purposes have not been, nor can they be, specifically determined at the time of the primary research, they can only be described in broad strokes or general terms in the initial informed consent process.

Researchers working for the sponsor to conduct secondary research analyses have no need, intent or reasonably available means to identify patients. Indeed, the purposes of secondary research typically are similar to retrospective epidemiological analyses and include, among other things, further analyses of factors involved in disease and treatment of disease. In both primary and secondary research using key-coded data, researchers within the sponsor organisation do not have access to the confidential key that would reveal data subjects' identities. Access to identified or identifiable information by field monitors and pharmacovigilance staff should not be imputed to sponsor researchers who use key-coded data for primary and secondary research purposes but do not have access to the confidential key.

The Working Party's Opinion already recognises that key-coded data that is considered personal data in the hands of one data controller may not be personal data in the hands of another. (See p. 20). What is critical is whether the data controller or any other person has means that are reasonably likely to be used to identify the subject of the information. Therefore, a data controller with access to the key might be considered to have personal data, while another person without such access may not have personal data subject to the requirements of the Directive.

The same can be true, however, of individuals within an organisation – there may be some individuals that have access to the key for specified purposes while others do not have such access. Where the key is maintained by a third party who is only permitted to allow access to the key on a limited basis and subject to stringent safeguards, it must be acknowledged that data that is identifiable in the hands of some individuals within the organisation may not be identifiable in the hands of others. This interpretation is consistent with the Working Party's conclusion that in the context of pseudonymised data, a "flexible" application of the data protection rules is justifiable. (See p. 18).

The IPPC is of the opinion that data protection rights are preserved in the secondary research context by instituting appropriate organisational and technical safeguards, as well as auditing mechanisms, to limit the availability of identified or identifiable data to those parts of a sponsor's organisation that need such access in order to perform their narrowly defined functions. Public interest supports allowing a scientific research organization to designate and segregate those parts that require access to identified or identifiable data from those that do not and applying the Data Protection Directive accordingly.

To require sponsors to obtain specific detailed consent for secondary research uses would necessitate recontacting subjects. Since subject contact information is held by external investigators involved in the original study, it is presumably these investigators who would need to contact patients, even though these investigators may not otherwise be involved in the secondary research project. The inability to recontact subjects (e.g., because of relocation of the data subjects or lack of cooperation by original study investigators who are not involved in the

secondary analyses) will reduce population sample size, thereby increasing statistical uncertainty in secondary research conclusions and in many cases completely prevent the research from proceeding. Moreover, in many cases the inability to recontact subjects to obtain consent may not be random and may vary in ways that bias study results.¹⁵ These risks and burdens are unnecessary given the protections already in place that prevent secondary researchers from identifying data subjects.

Biological Samples

Innovative healthcare research utilising genomics could be severely impacted by the Working Group's Opinion. The Opinion singles out human biological samples and genetic information derived therefrom as an area that creates unique data privacy considerations. The singling out of genetic information is inappropriate. Data derived from biological samples can be identifiable or non-identifiable, just as other personal health data can be. Genetic data are non-identifiable unless a reference database or similar available record source exists that links them to individual identities and this is accessible to the researcher.

Genomics involves the study of genes and gene products, such as RNA and proteins, and their respective functions. One particular sub-category of genomics is pharmacogenomics, which involves the study of the correlation between an individual's genotype and his or her response to drug treatment in order to deliver optimal therapies. Studying the genetic basis of patient drug responses enables pharmaceutical companies to develop safer and more effective treatments. By dividing large populations into "sub-populations" who share similar genetic characteristics, medicines can be developed that significantly reduce the risk of adverse events and/or increase the likelihood of a beneficial therapeutic outcome. In the future, pharmacogenetic testing will help facilitate the prescribing of optimal treatments to patients.

Many aspects of pharmaceutical development and monitoring are dependent upon the collection and analysis of a wide range of human biological samples associated with different diseases and treatments. In clinical trials, for example, the collection of blood samples is routinely undertaken as part of safety monitoring. In addition, drug regulatory agencies are increasingly looking to the industry and research institutions to employ innovative technologies, including genomics, to provide additional information pertinent to drug response.¹⁶

When collected, human biological samples may be obtained directly from volunteers or through a tissue repository in which samples have been stored for future research. At the time of sample collection, informed consent is sought to store the sample and use it for future research.

¹⁵ cf. S. J. Jacobsen et al., "Potential effect of authorization bias on medical record research", 74 *Mayo Clin. Proc.* 330-38 (1999).

¹⁶ See, *generally*, The European Medicines Agency Roadmap to 2010: Preparing the Ground for the Future (March 4, 2005) (EMEA/H/34163/03/Final).

The IPPC is concerned that the Opinion's conclusions about biometric data do not measurably increase privacy protections and would significantly hamper genomic research and other research involving human biological samples. The Opinion states that "Human tissue samples (like a blood sample) are themselves sources out of which biometric data are extracted. . . ." (p. 9.) The Opinion goes on to conclude that because biometric data "can be considered both as *content* of the information about a particular individual . . . as well as an element to establish a *link* between one piece of information and the individual" (p. 8), "the extraction of information from [tissue] samples is collection of personal data, to which the rules of the Directive apply." (p. 9.)

First, a minor point of clarification in terminology is appropriate to help explain the IPPC position. Simply because biometric data *may* be extracted from human tissue samples, it does not follow that all information derived from human tissue samples is biometric data. Biometrics is the science of using biological properties to identify individuals. Biometric data, in turn, is information that has been extracted from a biometric sample and is used either to build a reference template or to compare against a previously created reference template in order to identify individuals. Thus, in the phraseology of the Working Party, the term "biometric data" has both a *content* element – it references biological properties -- and a *purpose* element – it is used to identify individuals. Biological samples are commonly used in biomedical research, but it is important to understand that the purpose of such research most often does not include identification of the data subjects. The data derived from biological samples used in biomedical research, which does not involve an identified or identifiable biological sample or include identification of the data subjects, is not, therefore, "biometric data."

The second and larger point concerns the Working Party's conclusion that data derived from biological samples is necessarily identifiable. This conclusion appears inconsistent with the International Declaration on Human Genetic Data, which was adopted unanimously by UNESCO in 2003, and the Council of Europe 2006 Recommendation of the Committee of Ministers to member states on research on biological materials of human origin. Both of these documents recognize that data derived from human biological materials can be both identifiable and non-identifiable. In this regard, the Declaration defines three terms: "data linked to an identifiable person," "data unlinked to an identifiable person," and "data irretrievably unlinked to an identifiable person."¹⁷ Similarly, the Recommendation defines the terms "identifiable biological materials" and "non-identifiable biological materials."¹⁸

Both the Declaration and the Recommendation specify protections that are commensurate with the level of identifiability of the samples concerned. For example, the Declaration provides that a subject may withdraw consent to storage and use of genetic data "unless such data are irretrievably unlinked to an identifiable person."¹⁹ The Recommendation has a similar provision.²⁰ The Recommendation permits the use of unlinked anonymised

¹⁷ See Declaration, Art. 2 (ix) to (xi).

¹⁸ See Recommendation, Art. 3.

¹⁹ Declaration, Art. 9(a).

²⁰ Recommendation, Art. 15.

biological materials in research provided that such use does not violate any restrictions placed by the person concerned prior to the anonymisation of the materials.²¹ The Declaration also has a similar provision.²²

The Working Party appears to reach its conclusion that data derived from biological samples is necessarily identifiable because of an imprecise conception of what it means to “identify” an individual. The Working Party states that “a natural person can be considered as ‘identified’ when, within a group of persons, he or she is ‘distinguished’ from all other members of the group.” (p. 12.) He or she is “identifiable” when “although the person has not been identified yet, it is possible to do it.” (p. 12.) This interpretation is overly broad and could lead to impractical and futile regulation of data that cannot reasonably be linked to a specific individual. The concept of data individuation (*i.e.*, the differentiation of data as being related to separate individuals) must be distinguished from data subject identification (*i.e.*, discovery of the “identity” of the data subject).

We would assert that the concept of an “identifier” is more complex than simply a factor or set of factors that distinguishes one person from another. For example, information should not necessarily be considered “identified” or “identifiable” by virtue of the fact that unique, random numbers have been assigned to *distinguish* data about one person from another. A medical record that has been otherwise stripped of identifiers does not become identifiable merely because a unique, random code has subsequently been assigned to it. A DNA sequence, although potentially unique to an individual, should not be deemed to identify that individual in the absence of a database or similar available record source that links this sequence to the “identity” of the individual (*i.e.*, some socially consequential distinguishing characteristics, such as a name or contact information²³) and to which the researcher has access.

The Working Party seems to acknowledge this point of view in Example No. 13 in the Opinion. The IPPC agrees with the Opinion’s conclusion in Example No. 13 that the ability to distinguish pharmaceutical research data as having come from separate individuals is necessary but not sufficient to make that data “identifiable.” This is the essence of data individuation. In the example, although patient names are not used, “serial numbers [are] attributed randomly to each clinical case, in order to ensure coherence and to avoid confusion with information on different patients.” (pp. 15-16.) Nevertheless, despite the use of serial numbers to distinguish patients, the Opinion concludes that “a Data Protection Authority may consider that no means are present in the processing performed by the pharmaceutical company, which [*sic*] make it likely reasonably to be used to identify the data subjects.” (p. 16)

The concept of “identity” is highly relevant to interpretations of personal data in the context of human biological samples and many types of scientific research that distinguish data sets without necessarily identifying individuals. We therefore recommend that the Working

²¹ Recommendation, Art. 23.

²² See Declaration, Arts. 14 and 16.

²³ While the existence of such databases may hypothetically make the DNA information “identifiable” to those with access to such databases, it is important to take into account “all the means likely reasonably to be used either by the controller or by any other person”. The sponsor would not have access to such databases under any circumstances currently foreseeable.

Party further address the central issue of what it means to *identify* an individual and distinguish this concept from data individuation. The IPPC also recommends that in a future document, the Working Party, in consultation with other European institutions,²⁴ distinguish between data derived from biological samples that relates to an identified or identifiable person and data that does not.

EPIDEMIOLOGICAL RESEARCH

Epidemiological studies involve research on human populations in order to link human health effects to a cause. Certain types of epidemiological studies are conducted prospectively. For example, prospective cohort studies are one type of prospective epidemiological study which involves observation of a population over time to compare health impacts of one subset of individuals who were exposed to an intervention or other factor of interest to those who were not. Others types of epidemiological studies involve the retrospective review, collection and analysis of health-related information. One specific area of retrospective epidemiological research is health outcomes and cost-effectiveness research. Health outcomes and cost-effectiveness research evaluates and compares the costs and benefits of a particular pharmaceutical intervention in order to guide optimal allocation of healthcare resources.

In some types of epidemiological research, de-identified, aggregated data may be sufficient for researchers' needs. In still other research, such as longitudinal studies, data individuation such as that described in Example No. 13 may be necessary in order to monitor individual health outcomes over time. The IPPC supports the Working Party's conclusion in Example No. 13 that the random attribution of serial numbers to separate clinical cases, in order to ensure coherence and to avoid confusion with information on different patients, does not by itself render information identifiable.

Informational elements that could, in the absence of controls, be considered "indirect identifiers" are sometimes needed for epidemiological research. For example, dates related to an individual's treatment are often necessary to assess the sequence and timing of drug exposures and health-related impacts. Similarly, broad geographic identifiers are often needed (e.g., to compare the overall incidence of some condition in a population to those taking a preventive medication). The inclusion of this information should not be deemed to make data identifiable if an organization takes appropriate steps to prevent identification of the individual. As the Opinion states:

Putting in place the appropriate state-of-the-art technical and organizational measures to protect the data against identification may make the difference to consider that the persons are not identifiable, taking account of *all the means likely reasonably to be used by the controller or by any other person* to identify the individuals. In this case, the implementation of those measures are not the *consequence* of a legal obligation arising from Article 17

²⁴ For example, it would be appropriate to consult with the EMEA, who has been involved in the development of ICH E-15 on Terminology in Pharmacogenomics.

of the Directive (which only applies if the information is personal data in the first place), but rather a *condition* for the information precisely not to be considered to be personal data and its processing not to be subject to the Directive.²⁵

The IPPC supports this interpretation.

CONCLUSION

The IPPC appreciates this opportunity to comment on the Working Party's Opinion and the related issues that it raises. We look forward to engaging in a dialogue with the Working Party as its work concerning pharmaceutical research and regulatory compliance advances.

DC01/ 532620.6

²⁵ Opinion at 17.