

Privacy issues in pharmacoepidemiology: the importance of weighing costs and benefits

Hubert G. Leufkens PhD*

Department of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences (UIPS), The Netherlands

KEY WORDS — record linkage; medical data; privacy; ethics; The Netherlands

INTRODUCTION

As the availability of both drug exposure and outcome information becomes increasingly important to pharmacoepidemiology, the ability to link person-specific clinical, drug use and disease course data is a critical objective. Based on the Declarations of Helsinki from the World Medical Association (see for most recent version <http://www.wma.net/e/policy/17-ce.html>), it is a basic right of the patient to be assured that all his medical and personal data are confidential.¹ Only in the case of a few well-defined exceptions is disclosure of person-specific information allowed, e.g. prevention of serious risk to public health, order by a court of law in a criminal case, and under certain safeguards, scientific research. The tension between assuring personal privacy and access to medical data for epidemiological research has drawn ample attention from various stakeholders (individual patients, the public, politicians, health professionals, and the research community). In the advent of a surge in health care databases, and computerization of medical practices, many countries both in North America and Europe have taken comprehensive legal action to assure the protection of personal privacy.^{2–4}

The scientific community usually responds according to two scenarios to possible threats of restrictive legislation: the route of political action and the pragmatic/technological route. The first aims to convince decision-makers of the importance of collective benefit to society from research with medical data and that we cannot rule out significant adverse effects to public health when epidemiological research has been made virtually impossible.^{4,5}

The pragmatic route takes a technological departure using methodology that includes contemporary computer and statistical technology in order to build, within the framework of existing privacy legislation, aggregated, de-identified but person-specific, information.^{6–8} Both scenarios have their own merits and should continue to contribute to finding solutions that serve both public and personal interest in privacy protection and scientific needs of using critical health data.

EXPERIENCES WITH RECORD LINKAGE IN DUTCH PHARMACOEPIDEMIOLOGY

In this section we review briefly a number of experiences of record-linkage for pharmacoepidemiology in the Netherlands. As in most countries, the Dutch experiences represent a patchwork of various approaches to link individual sets of drug exposure and clinical data. In the absence of a (national) unique identifier, researchers have to rely on other approaches for bringing separate datasets together to a patient-specific linked set. Essentially, four different strategies have been applied.

*Correspondence to: Dr H. G. Leufkens PhD, Department of Pharmacoepidemiology and Pharmacotherapy, UIPS, PO Box 80.082, 3508 TB Utrecht, The Netherlands. Tel: +3130-2537323. Telefax: +3130-2539166. E-mail: H.G.M.Leufkens@pharm.uu.nl

Firstly, datasets have been linked by using a clearing-house model. In such a model a 'neutral entity' is allowed to use personal information for linking two or more datasets to an anonymous, de-identified set of data. The 'neutral entity' can be an individual or a dedicated institute with the special task and obligation of data linkage without any disclosure of the personal data items necessary for the linkage procedure. An example where the pharmacists had the role of the 'neutral entity' was a study where data of patients using NSAIDs heavily, were linked with information from GPs on the prevalence of arthritis and other rheumatic diseases.⁹ In order to prevent disclosure of medical data to the pharmacist in his role as 'neutral entity', a procedure with sealed envelopes was applied for transporting the queries for detailed clinical information on the patient's diagnosis and disease course from the investigator to the GP via the pharmacist. The investigator using anonymous pharmacy records asked the pharmacist, who was in the position to link these records to the name of the patient and his/her GP, to send the query to the GP with the request to complete a listing of questions on that individual patient. The completed questionnaire was then sent back to the pharmacist in a coded sealed envelope, enabling the contents of this envelope to be looped back to the investigator without violating the confidentiality of the individual patient. The pharmacist operated in a situation which allowed him/her to link two confidential sets of data—prescription records and clinical data from the GP—in an anonymous way for the investigator. Although this was a feasible way of record linkage, it was very time-consuming, costly and not very applicable for routine use.

Record linkage in Dutch pharmacoepidemiology became mature in the early 1990s when Herings *et al.* developed PHARMO.¹⁰ This system links community pharmacy and hospital data within established hospital catchment regions, on the basis of patients' birth date, gender, and GP code. The linkage is based on a probabilistic model of the combination of these three data items yielding a sensitivity and specificity of over 95% each. The PHARMO system has been expanded to a population of over 500 000, is population based, and links all prescription drug data to hospital data. Presently, PHARMO is also being linked to primary care data, population surveys, cancer and accident registries, and other outcomes data. The data collection is longitudinal and goes back to 1987. So far, data from the PHARMO system have been used in more than 50 studies. The system has also provided a productive model for linking other health care data sets for use in Dutch epidemiology.

As stated before, the Dutch health care system does not have a unique patient identifier, although some regional hospitals are developing such codes in order to facilitate seamless care for patients being hospitalized. In the Tilburg region (south west of the Netherlands), all patients frequenting the local hospitals, GP practices, pharmacies and other health care facilities do have one unique personal code. Regional patient identifiers are still in an experimental phase and certainly not universal. In a study on the channeling of new antidepressants, data on drug use at home was drawn for community pharmacies and linked to all other relevant patient data for the study through this regionally unique number.¹¹ It is anticipated that the Dutch health care system will be organized more along regional structures in the future. An important driver for this development is the fact that the Dutch government is promoting the role of regional health care insurers as main stakeholders for planning, financing and policy making regarding the health system. Therefore it is expected that this model of regionally unique patient numbers will be more prevalent in the next decade, and probably a useful tool for epidemiological research.

Finally, the fourth approach is to acquire informed consent from each individual patient. This strategy was for instance applied in a follow-up postmarketing surveillance study on a new proton pump inhibitor.¹² Less than 2% of the pre-screened patients eligible for inclusion in the study refused informed consent. All data and documents related to patients were kept in strict confidence and in accordance with the official privacy regulations. Data were collected at the visit of inclusion in the study and at each follow-up visit during lansoprazole therapy. There was a maximum follow-up of 2 years, performed by reviewing the medical file and by patient-questionnaire. The data collection aimed not to influence normal procedures in any way. Medical data were recorded by participating physicians while pharmacists provided drug-dispensing records related to the study patients. While this model requires a prospective study design, it remains vulnerable to selective refusal of informed consent by relevant patient groups. This was probably not the case in the lansoprazole study (only 2% refusal rate), but could be a problem in studies where drug use has a more emotional context (antidepressants, cancer drugs).

Dutch privacy regulations include basically the law on person registries and the law on individual health care agreement. The first law is a general law applicable to all areas in society and is about how to deal with confidentiality and privacy issues when

personalized data are kept in a registry and is in line with EU regulations. The law provides a legal framework for the practical and organizational conditions under which a registry can be maintained, who has access to what data, consent considerations and the requirement that all person registries should be catalogued in a national chamber of all registries. The law on individual health care agreement regulates the relationship between the patient and the practicing health care provider (physician, nurse, and hospital pharmacist) and how within the context of individual health care provision clinical data are kept. Both legal frameworks together need to be considered in case one wants to use patient data for research purposes.

BEYOND THE PRAGMATIC SCENARIO: WEIGHING ETHICAL 'BENEFITS' AND 'COSTS'

Experiences in coping with requirements to assure data privacy in the pragmatic scenario have been dominated mainly by technical (e.g. probabilistic linking, de-identification, pseudonymization, introduction of random error on an individual level but not on a population level, etc.) or procedural (e.g. standard operating procedures, good practice standards, security, etc.) dimensions. In the background of all this, ethical principles are there, but in most cases virtually invisible. The thesis of this short report is that weighing of ethical 'benefits' and 'cost' is an essential and productive additional perspective in designing and conducting sound pharmacoepidemiological research. Ethical reasoning helps also to be concise in defining the research question, the design and the conduct of the study. It is not necessarily a cumbersome 'burden'.

The ethical perspective deals also with the data management and disclosure approaches in balancing the interests of science while respecting confidentiality rules and guidelines. Assuring data privacy and confidentiality of persons has its origins in the fulfillment of relevant ethical principles. In general, four ethical principles can be distinguished.¹³

Beneficence

Mankind has a moral obligation to do 'right', i.e. to benefit individuals and society. For scientific work this means that a research project should add to the existing knowledge base of medicine in order to improve the health of patients, to prevent health hazards and mortality.

Non-maleficence

This principle reflects a moral obligation not to do harm to the persons involved in a scientific study. Under certain circumstances harm can be justified when the population benefits outweigh the individual harm.

Autonomy

The principle of autonomy is closely linked with the right to self-determination. Autonomy is the key principle for respecting personal privacy. In clinical research this principle is reflected in the asking for informed consent when a patient enters a study.

Justice

The principle of a fair distribution of burdens and benefits between people is an important feature of most modern societies. This principle may mean equal 'risk' in terms of drug exposure, for instance in post-marketing research when there is still uncertainty about certain drug effects.

Nilstun and Westrin have proposed to apply these principles from the perspective of each of the parties involved and then to assess the ethical 'benefits' and 'costs' in the event that the study is or is not conducted.¹⁴ In this short report we apply this model to a recently conducted follow-up study on the differential risk of venous thrombosis (VTE) in second and third generation oral contraceptives (OCs).¹⁵ The model starts with identifying the relevant parties (females using OC, society at large). In Table 1, a possible outcome of an analysis of the most relevant 'benefits' and 'costs' is listed concerning the two dimensions of ethical principles and parties involved in the event that the study is conducted. If the study is carried out, there are possible 'benefits' for society at large, in particular (other) women using OCs in this case, because the study confirmed earlier findings that most of the VTE risk is concentrated in the very young

Table 1. The most important possible 'benefits' and 'costs' when the study is carried out¹⁵

| | Beneficence | Non-maleficence | Autonomy | Justice |
|----------------------|-------------|-----------------|----------|---------|
| Females in the study | | | Costs | |
| Society at large | Benefit | | | Benefit |

users of OC. This has provided important guidance to decision-makers and young females in choosing the most suitable OC. There could be potential costs with respect to autonomy by violating the privacy of the women in the study. However, all data used in the study were anonymous, meaning that the investigator could not link research data to any individual women. The linkage procedure of PHARMO has been submitted to and approved by the Dutch National Chamber of Person Registries. There is a development in the current privacy debate in the Netherlands to ask for general informed consent to use medical data for research purposes at the time a patient enter the health care system. This would allow using the data in future studies without asking for consent for each individual study. What the outcome of this debate will be in terms of practical solutions is still unclear. Of course, serious assessment of each individual research protocol by an independent review board will be a mainstay in such a system.

From the perspective of 'justice' we may add that by 'participating' in such a study the women involved took their share in the solidarity of bringing together relevant data for solving an important public health problem. However, one should point out that 'participating' was not a conscious act for the individual women whose data were used in this study.

CONCLUSION

Computerization of databases has increased apprehension about loss of privacy. Record linkage is a key methodology in clinical research and epidemiology and various techniques have been developed to cope with the confidentiality issue.^{7,16} The creation of unbiased personal histories (including both data on various exposure and outcomes) is a crucial requirement in pharmacoepidemiology.^{9,10} Ethical weighing of 'benefits' and 'costs' can play an additional and relevant role as a vehicle for thoughtful reasoning. Indeed, concerns have been expressed about the various ways of misusing such data.¹⁷ In particular in the era of genetics and the increased interests of health insurers to reduce their business risks, there is a great need for prudence, protection and careful weighing.^{2,5} At the end of the day, however, we all are losers when the research community,

clinical medicine and patients are not able to solve this in a way that provides mutual benefit for all stakeholders.

REFERENCES

1. Beauchamp TL, Childres JF. *Principles of Biomedical Ethics* (3rd edn). Oxford University Press: New York; 1989.
2. Gostin L. Health care information and the protection of personal privacy: ethical and legal considerations. *Ann Intern Med* 1997; **127**: 683–690.
3. Tondel M, Axelson O. Concerns about privacy in research may be exaggerated. *Br Med J* 1999; **319**: 706–707.
4. Vandenbroucke JP. Privacy, confidentiality and epidemiology: the Dutch ordeal. *Int J Epidemiol* 1992; **21**: 825–826.
5. Willison DJ. Health services research and personal health information: privacy concerns, new legislation, and beyond. *CMAJ* 1998; **159**: 1378–1380.
6. Subramanyan GS, Yokoe DS, Sharnprapai S, Tang Y, Platt R. An algorithm to match registries with minimal disclosure of individual identities. *Public Health Rep* 1999; **114**: 91–93.
7. Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, Dusserre L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. *Methods Inf Med* 1998; **37**: 271–277.
8. Blakely T, Woodward A, Salmond C. Anonymous linkage of New Zealand mortality and Census data. *Aust NZ J Public Health* 2000; **24**: 92–95.
9. Leufkens HG, Ruter EM, Ameling CB, Hekster YA, Bakker A. Linkage of pharmacy data on heavy users of nonsteroidal anti-inflammatory drugs to information from general practitioners. *J Pharmacoepidemiol* 1991; **2**: 67–77.
10. Herings RMC, Stricker BHCh, Nap G, Bakker A. Pharmacomorbidity linkage: a feasibility study comparing morbidity in two pharmacy-based exposure cohorts. *J Epid Comm Health* 1992; **46**: 136–140.
11. Leufkens H, Claessens A, Heerdink E, Eijk J van, Lamers CBHW. A prospective follow-up study in 5,669 users of lansoprazole in daily practice. *Aliment Pharmacol Ther* 1997; **11**: 887–897.
12. Egberts AC, Lenderink AW, de Koning FH, Leufkens HG. Channeling of three newly introduced antidepressants to patients not responding satisfactorily to previous treatment. *J Clin Psychopharmacol* 1997; **17**: 149–155.
13. Gillon R. Medical ethics: four principles plus attention to scope. *Br Med J* 1994; **309**: 184–188.
14. Nilstun T, Westrin CG. Analysing ethics. *Health Care Analysis* 1994; **2**: 43–46.
15. Herings RM, Urquhart J, Leufkens HG. Venous thromboembolism among new users of different oral contraceptives. *Lancet* 1999; **354**: 127–128.
16. Roos LL, Nicol JP. A research registry: uses, development, accuracy. *J Clin Epidemiol* 1999; **52**: 39–47.
17. Knox EG. Confidential medical records and epidemiologic research. *Br Med J* 1992; **304**: 727–728.