Med Law consult

Patient data for health research

A discussion paper on anonymisation procedures for the use of patient data for health research

incorporating:
Zorggegevens voor zorgonderzoek
De TTP als panacee?

mr. Evert-Ben van Veen

October 2011

The Hague, October 2011
ISBN 978-90-75941-00-5

First print, 2011

Printed copies can be obtained from MedLawconsult at Euro 15.00 each including postage and packaging.

Layout:
OGGI and Graham Kennett (MedLawconsult)

PREFACE

What started as a simple inquiry into the possible role of a Trusted Third Party (TTP) for health research in the Netherlands, became much more. Through a TTP patient data from various sources could arrive at the research domain under a unique pseudonym but anonymously for the researchers concerned. Researchers would have all the data they need, the privacy of patients would be protected and the tension between privacy and research for the public good would resolved. However, during the inquiry it proved that this 'best of both worlds' does not exist. The main problem is that epidemiological researchers in particular often need data on such a detailed level that they cannot be considered anonymous according to the very strict standards for anonymous data following from the interpretation of Directive 95/46/EC.
This lead to a further inquiry. The emphasis on a TTP can be seen as a lack of trust, namely that indirectly identifiable data would be identified by researchers. It is also an example of regulation by technology, trying to avoid a moral problem. A description of the role of a TTP for health research wouldn't be complete without addressing those additional issues. As researchers outside the Netherlands became interested in those, these issues are discussed in English.

This Report therefore consists of two Parts. The first part is the full inquiry into the role of a TTP in the Dutch debate of patient data for health research. That Part is in Dutch and summarised in English (infra 1-10). The second Part of this Report is written in English. A practical solution is proposed for the challenge that data will be will always identifiable in the research domain (Chapter 9). In Chapter 10 I discuss the moral or ethical aspects of using patient data for health research. The concluding Chapter 11 discusses some of the remaining questions if the practical solution would be implemented, and makes some final comments.
This set up means that there is partially an overlap between the Dutch and the English Parts. I have tried to avoid that overlap in the summary.

Working on this Report has not always been easy. In the Dutch part the present state of the law is described as accurately as possible but especially chapter 10 is 'against the current' of the prevailing legal and regulatory doctrine. At the same time it could also be argued that it shows an 'undercurrent' of thought which deserved to be expressed. But one has to be twice as cautious then.
Many have put their trust in me that I could finish it. The FEDERA and its indefatigable chairman Jan Willem Coebergh who commissioned the Dutch part of this inquiry, researchers and representatives of patient organisations who gave valuable comments on earlier drafts, EUROCOURSE which made funds available to contribute to the English chapters and friends such as Pieter Ippel and Leigh Hancher who commented encouragingly on a paper which ultimately became Chapter 10. My colleague Graham Kennett who corrected the English and other errors in several earlier versions. I thank them all. Yet the text and all possible mistakes are my own responsibility.

Evert-Ben van Veen
October 2011

INHOUDSOPGAVE

LIST OF ABBREVIATIONS

AWBZ	Algemene wet bijzondere ziektekosten
BVerfG	Bundesverfassungsgericht
CBP	College Bescherming Persoonsgegevens
CNIL	Commission Nationale de l'Informatique et des Libertés
DPA	Data Protection Authority
ECHR	European Court of Human Rights
EHCR	electronic health care record
FEDERA	Federatie van Medisch Wetenschappelijke Verenigingen/
GCP	Good Clinical Practice
HIV	Human immunedeficiency virus
ICT	information communication technology
METC	Medisch-ethische toetsingscommissie
NAW	Naam, adres, woonplaats
NHS	National Health Service
NZa	Nederlandse Zorgautoriteit
PET	Privacy Enhancing Technologies
TFEU	Treaty on the functioning of the European Union
TTP	Trusted Third Party
Wbp	Wet bescherming persoonsgegevens
WMG	Wet marktordening gezondheidszorg
Zvw	Zorgverzekeringswet

ENGLISH SUMMARY

ENGLISH SUMMARY

FROM THE DUTCH PART

1.
Epidemiological research or health research in general is impossible without using patient data. The use of such data by researchers is subject to the privacy protection of the patients concerned. In principle the patient should have given explicit consent that personal data concerning his or her health can be used by researchers.
Patient data are here considered as a specific kind of personal data. They are 'sensitive data' in the sense of Directive 95/46 EC. The opposite of personal data, and hence patient data, is anonymous data. Anonymous data fall outside the scope of the Directive and anonymised patient data can be used for research without consent. The Directive determines when data are anonymous or should be considered personal data.

2.
Fully anonymous data are more often than not, unsuitable for health research. The more intricate relations between exposure, the onset of disease, treatment regime and/or results of treatment can usually not be investigated with fully anonymous data.

Directive 95/46/EC provides for exemptions to the consent principle for the purposes of health research in the general interest.
The privacy legislation of (nearly) all EU countries have implemented such exemptions, though each in a very different way. According to Dutch law, namely the Act on the Treatment Contract (article 7:458 BW) and the Data Protection Act (article 23.2) combined, patient data can be used for health research without consent in two instances:

a. It is not reasonably possible to ask for consent and the privacy of the patient is not unnecessarily jeopardised.
b. Given the nature of the research, asking for consent is not feasible and the data arrive at the researcher in such a way that re-identification is sufficiently prevented.

In both instances three other conditions have to be met, being:

- The research serves a general interest;
- The research cannot be carried out without those data;
- The patient has not objected to such use of his or her data for research.

The first instance is, amongst other things, applied to situations when the patient is deceased or, also after a reminder, does not respond to a written request asking for permission to use his or her data for a specific research.
The second instance applies to situations where many patients (more than a few thousand) would retroactively have to be asked for consent.
In that first instance directly identifiable patient data may also be used. In the second instance only indirectly identifiable data may be used.
The rather general provisions of Dutch law have been elaborated in a Code of Conduct by the FEDERA.

3.
The problems with explicit consent for the use of patient data for health research are often underestimated.
The interaction between doctor and patient in the consulting room is geared to something completely different. Namely to address the concerns of the patient about his or her condition and for the

doctor to acquire, with informed consent, data necessary to arrive at a diagnosis and treatment plan, the treatment, readjustment of the treatment plan, etc.
The early phases of medical interventions for serious conditions leaves no room to discuss something not connected with that primary aim of the interaction in the consulting room such as using those data for later health research. For patients where 'disease management' has become more or less stable this might be different. While explicit consent for 'further use' of data is already quite difficult (against a background of taking informed consent about the patient's course of treatment very seriously) it cannot be true informed consent as it is usually unknown at that time for what specific protocols the data will be used. For such informed consent all patients would have to be re-contacted, insofar as still possible, often years after the treatment has ended.
The second problem is that consent is shown to create bias, especially when asked for retrospectively, where non-consenters are often the socio-economical more vulnerable groups.

4.
The report starts with the presumption of a 'chain of data'. Direct access to health care files (or records) is not considered in the Report since a TTP can not have a role in such a case. The 'chain of data' means that there are data at the source assembled in the context of the function of that source, such as data assembled by a health care provider for the treatment of the patient, and then these data are sent to a receiver, in this case the researcher. Often it is necessary for the research project to receive data from other sources as well.
The report states that researchers are not interested in persons as such but in patterns about persons. Data relating to one unique person must be gathered, as has been said usually from different sources, to investigate – with sufficient statistical power – the kind of relations mentioned above. The researcher does not need to know who that person is, but needs to be assured that data relate indeed to one specific person and not to someone else, in order to acquire reliable data. Hence the maxim throughout this Report: 'patterns, not persons'.

In connection with the consent problems mentioned, the use of TTP is seen as the solution to the privacy versus research dilemma. In between source and receiver the TTP uses techniques to pseudonymise the data. Through a TTP the researcher would only receive anonymous data in the sense of Directive 95/46/EC. If data are used from various sources, the TTP construction would guarantee that data from those sources reaching the research domain uniquely relate to the individuals at the source, yet are still anonymous.

5.
It is hard to find a unique definition of a TTP. *The* standard TTP does not exist. The TTP originated in electronic commercial transactions as an independent, reliable party in-between the sender and the receiver of data, usually to ascertain the authenticity of the sender and/or of the data. In health care the TTP functions to *mask* the identity of the data subject. The patient data will be pseudonymised. In Dutch usually the phrase 'coded' is used for such data.

6.
Some authors in the Dutch medical law discourse claim that pseudonymised or coded data always would remain identifiable patient data in the sense of the Act on the Treatment Contract.
That position is mistaken. Identifiable patient data are a species of personal data in the sense of Directive 95/46/EC. When coded, the answer to the question whether or not they can be considered anonymous, is the same as with other personal data. That answer will depend on:
- the safety of the pseudonymisation procedure: can the pseudonym in one way or the other be traced back by the receiver at the end of the data chain to the identity of the data subject;

- the aggregation level of the research data attached to the pseudonym (or code-number in the Dutch discourse). In other words: are those research data not indirectly identifiable.

With the latter the following is meant. Data on which research is based (research data) need always pertain to a specific person, which can be discerned from other persons in the research database. Age group, gender, onset of disease, details about the disease, sometimes profession and location are often relevant research data but can under circumstances be considered indirectly identifiable. The aggregation level refers to the degree of specification of these data, pertaining to one individual. Certain classes of similar data can be grouped together. A higher aggregation level means that more data of that class have been grouped together, so instead of age by year, '5 year age groups'. Or instead of the full zip code, only 3 digits (the Dutch zip code has 6).

Hence pseudonymised data can be considered anonymous data, even if the coding is reversible at the source (in the Dutch debate we speak of 'two way coding'). Opinion 4/2007 on the concept of personal data by the article 29 Working Party recognizes this logic of the data chain. Data may identifiable at one point in the data chain, but not at a later point which only holds pseudonymised data. The Opinion discusses the conditions which have to met in such a case. The conditions are quite elaborate, to avoid unwanted, and illicit, 'replay back' of the pseudonymisation.

7
This leads to the following scheme of data in the sense of Directive 95/46/EC

<table>
<tr><td rowspan="2">Anonymous data</td><td colspan="2">Fully anonymous data</td></tr>
<tr><td colspan="2">Coded anonymous (pseudonymised) data</td></tr>
<tr><td rowspan="3">Personal data</td><td rowspan="2">Indirectly identifiable data</td><td>Coded but either coding insufficiently secure or aggregation level too low</td></tr>
<tr><td>Not coded but aggregation level too low</td></tr>
<tr><td colspan="2">Directly identifiable data</td></tr>
</table>

8.
The TTP functions already in Dutch health care, both in the context of the health care system as well as in health research. In the context of the health care system a TTP is, amongst other things, used to acquire reliable data about waiting lists of mentally handicapped people whose guardians have applied for long term care. As guardians tend to apply to more than one health care provider, the combined waiting lists of health care providers did not give accurate data of the number of people waiting for a place. A system was set up by which the waiting lists of the respective health care providers would be combined. It should be an anonymous list in order to have accurate data for policy decisions about allocation of funds. The list would be kept by an Agency which has a prominent role in the Dutch health care insurance system.
This case is the only example where the Dutch DPA has issued a public statement about the conditions for pseudonymisation through a TTP. They are as follows:

a. pseudonymisation is done professionally. The first encryption should be performed at the source;
b. technical and procedural measures have been taken to prevent re-identification through the encryption ('replay back');
c. the aggregation level of the data will not make them indirectly identifiable;
d. an independent audit at the start of the process should ascertain that the conditions mentioned above have been fulfilled. This audit should be repeated regularly;
e. the pseudonymisation process should be clearly described in a public document, so that the data subject can check the measures taken.

As will be shown, those conditions are hard to meet in much of today's health research.

9.

The TTP in health research works in a similar way. It is an independent and reliable party which takes care of the controllable pseudonymisation of patient data into anonymous data as described above.

Contrary to what seems to be suggested by ISO 2008 the TTP does not receive personal data to convert them into pseudonymised anonymous data. From a privacy perspective nothing would be won then. In short, the TTP takes care of the techniques by which the source can convert those data into the aforementioned pseudonymised anonymous data. In reality it is a bit more complicated but this is the basis.

10.

As mentioned, a TTP seems to have many advantages. Data from different sources can be uniquely (to one person) combined. Yet that person remains anonymous to the receiver of the data. There is – or seems to be – no privacy issue and the restrictions on the use of patient data for health research don't apply.

However, a TTP construction has some major disadvantages as well. From the perspective of the researcher these are:

- The construction is rather costly. It can be a cost efficient way to link large datasets from various sources, but such research will need additional funding. Many small-scale research projects, funded by private foundations set up to support patients suffering from a specific condition, would lack the means to employ a TTP.
- The data need to be anonymous regarding the aggregation level of the data relating to one person and hence also the combined set of data under a certain pseudonym. Such truly anonymous data are often too general to investigate more refined correlations between exposure, onset of disease, treatment and/or results and to eliminate possible confounding factors, especially in rare diseases.
- This will be even more so if fundamental challenges made by statisticians are taken seriously. Given the amount of data which nowadays freely float around on the internet, and with which batches of not directly identifiable data can be combined, it has become very difficult to speak of anonymous data.
- The TTP applies 'one way coding'. This means that also the TTP cannot decode the pseudonym. However, researchers often need more details about one or more specific persons. For example if the record submitted to them shows inconsistencies such as when the medication cannot correspond to the diagnosis. They need to check this with the submitting physician. It should not be underestimated how much of this work consists first of all of carefully scrutinising and analysing data on their accuracy. Much more a 'métier' than large computional throughput of data, though that will come later with data which have been checked to be correct in earlier phases.

- In larger longitudinal studies, the pseudonym is at a certain moment replaced by a new pseudonym, in order to avoid 'replay back'. The old dataset cannot be used anymore. But these old data are often still necessary, first of all as analysing the data might take years and secondly to answer challenges by other researchers of the publicised findings based on the old data.
- Hence, 'the best of both worlds', being both anonymous data and proper health research with data does (to a large extent) not exist.

The last disadvantage is of a different nature. The TTP is fundamentally an answer to distrust – not an answer to a privacy problem. If research data reach the researcher indirectly identifiable, either because of a primitive pseudonymisation procedure or because of the aggregation level or a combination of both, they are considered personal data as the presumption is that researchers will try to retrieve the identity of the subject. The Report gives some examples where the Dutch DPA ruled from this presumption. Researchers were insulted by this presumption, yet this could only be expressed in a very diplomatic way, as the researchers needed the DPA to get some kind of approval, in this case using the research exemptions in Dutch data protection law.

11.

Within the present Dutch law the following can be concluded:

- The TTP should be seen as the ultimate PET (privacy enhancing technology) in health research.
- PET are then seen as technologies to mask the identity of the data subject as much as possible for the user of those data at a later point in the data chain insofar as this does not endanger the functionality of the data chain.
- PET should always be used in health research. An example, already in use in all such known studies and an element of the FEDERA Code of Conduct (2004), is that in cohort studies based on questionnaires, the department sending out the questionnaires and having the addresses, will be different from the department with the researchers receiving the answered questionnaires.
- The use of TTP as a way to send pseudonymised data to the research database should be based on various factors, such as:
    - Whether it is necessary to link various source databases;
    - Whether there is no explicit consent to use these data or there are flaws in that consent (such as too specific and not covering the present project);
    - Whether the aggregation level of the data when using a TTP still allows for research into sensitive correlations. As noted, often it does not.
- If those conditions do not apply, it might still be possible to use a TTP. However, the TTP will in that case not be used to make data anonymous but simply that indirectly identifiable data from various sources arrive at the researcher under an unique pseudonym and hence can be linked. The TTP might have more cost efficient techniques for such linking than when a researcher would to try create a PET so that data will arrive in the research domain under a pseudonym.
- However, that will not be a TTP function in the original sense.
- The same applies if the researcher would need to inquire more about the data provided (as seen in the examples at point 10) this TTP in the original sense will not work as such inquiries are impossible due to 'one way coding' system at present.
  Hence the conclusion is that a TTP in the original sense, namely to completely mask the identity of patients, can only be used in a rather limited realm of health research or more specifically of epidemiological research. The TTP construction cannot make the present research exemptions in Dutch law obsolete. In many instances researchers will still need them.

FROM THE ENGLISH PART

12.
The challenge that when data in the research domain *can* be re-identified, they *will* be re-identified, has lead to the search for the holy grail of anonymous data. As indicated already, this is a fruitless path.
It is fruitless for two reasons. Researchers often need data which are indirectly identifiable by present standards. These standards might even be considered obsolete as the new maxim by some authors is that nearly all supposedly anonymous data can be combined with today's wealth of personal information on the internet and hence be re-identified. Anonymous data may have ceased to exist.
The quest for anonymous data would lead research into a data desert. The focus of regulatory oversight should instead shift from establishing whether data are anonymous *before* they reach the research domain to the safeguards *within* the research domain that the persons behind the data with not be illegitimately re-identified. As said: researchers are after patterns, not persons.

13.
Regarding the data chain of research data we have a much less complicated situation than with the endless dataflow over the worldwide web, social networks, cloud computing and the like. Those create huge challenges to the present regulatory regime of data protection. The data chain for research data can be a closed one. At all points in time it can be ascertained which data have been sent to which recipient and who is responsible for them. The challenge is then to lay down the conditions under which the data will handled by the recipient (the research domain) according to the terms by which they were sent, and see that they will not be re-identified.

14.
The conditions for considering data in the research domain secure against re-identification cannot be merely technical. Establishing them should start with the recognition that researchers have one vested interest, which differs from commercial entities holding data. The very essence of the researchers business is those data, their integrity and their confidentiality and compliance with the trust with which sources have submitted data. If that trust would be violated, this would mean the end of the data chain and most probably of the researchers career. Putting all controllers of data, whatever their background and interests, on equal footing denies this fundamental difference and is partially the cause of the present problems. It denies, amongst other things, the professionalism of researchers.

15.
Trust must also be earned by accountable and transparent procedures. Researchers must show how:

- The data chain is safe against illegal access by third parties;
- Data will only be used according to the research protocol for which data were submitted by sources (whilst allowing for amendments if the inherent serendipity of research with data would take a different turn);
- Hence the data subjects behind – possibly – indirect identifiable data, will not be re-identified within the research domain either.

Appendix 1 describes in general which procedures are appropriate in this respect and how it can be ascertained by regulatory authorities whether they are in place.

16.
This would correspond with the meaning of identifiable data in Directive 95/46/EC, holding the phrase 'means likely reasonably to be used' (in Recital 26). The conditions are then about when re-identification is not reasonably likely.

17.
The above seems a technical–legal discussion about the conditions for using patient data for health research. That technical discussion must be seen against a background of viewpoints which consider the use of patient data for health research as exceptional. Such exceptional use can according to those viewpoints jeopardise the interests of the patient and certainly his 'autonomy'. Hence it should be avoided as much as possible.
Chapter 10 challenges those views which prevail in medical law but which seem to be less endorsed in recent literature on (public) health ethics.
In a way the prevailing view is understandable. Patient data are first of all collected for the treatment of the patient concerned. Medical law has greatly empowered the patient in the treatment relation and nearly all aspects of informed consent are applied correspondingly to the use of the data which are necessarily assembled about the patient during his treatment to assist that treatment. Such assembling is the *individual* function of patient data. The recent trend to view the patient more as a consumer of health care services adds more empowerment to patients.

Yet this consumerism has its limits in the European solidarity based health care systems. Health care delivery is based on statutory entitlements which are need and not demand oriented. Within such systems patient data have in various applications and using appropriate PET, a *collective* function as well. That collective function is equally as important as the individual function. Without that function equitable health care systems could not exist. The collective function should not be considered as an exception to the use of patient data as is done in most European documents.

18.
Health research can be seen as one of those collective functions and is also an example of the limits of consumerism in health care. All people submit at least some data when applying for a product or service. When a patient does so, those data are very detailed and intimate and often only become available *through* the service, namely as results of diagnostic procedures. They were unknown to anyone, including the 'consumer', beforehand. However, the major difference between patients who submit data for a product or service and all other consumers is the following: the patient does not only use health care but is in his or her very essence, health, wellbeing, life and death, the "*end product*" of that service as well.

19.
Without using patient data, the health care of that "end product", the patient itself, can never be improved. Patients expect health care to be optimal. When applying for the best possible care, patients are dependent on the use of data of many patients treated before them, which made this health care possible according to the best present standards.

20.
In many jurisdictions the patient is granted nearly full control about how his or her data will be used for the individual function. In that case only that patient will suffer the consequences if these

decisions were not very sensible according to the standards of most others. In the collective function, others will suffer the consequences if the data can be solely used according to the preferences of the patient concerned.

21.
Hence, the discussion should not be seen as a tension between 'privacy' on the one hand 'freedom of research' on the other hand but between informational self-determination on the one hand and on the other hand the right to the best possible care of other patients or the prevention of risks for the population at large.

22.
The discussion might be different if not all patients would have equal access to the same health care and the innovations through research, when validated and assessed on a cost/benefit ratio, embedded in that. But in European health care systems all patients have such equal access. So all (future) patients will equally profit from the results of health research. The moral challenge is why we have solidarity between the healthy and the sick, between the wealthy and the poor in that system and not between those whose data can be used and those who will profit from those data by better health care.

23.
The usual answer is that the patient's autonomy is involved here. Two issues are easily confounded in this context. The aim of health care and health care providers is to restore the patient's autonomy in the face of disease. Informed consent enabling the patient to decide about diagnostic procedures and treatment is an essential part of this. Deciding about data for the individual function can be so as well. Deciding about data for the collective function is not about this restoration of autonomy in the face of disease but concerns autonomy as informational self determination. That informational self determination can be trumped by the public good if certain conditions are met.

24.
Limiting the use of data to the disease for which the patient was treated, denies the huge interdependency between diseases, their underlying causes, and how medical interventions (by pharma, by diagnostic procedures, by training etc.) have been developed.

25.
The use of patient data for the various applications of the collective function should always be strictly necessary and proportional to the specific aim. PET should be employed to achieve that. Monitoring of the health care system can be done with aggregate data on a high aggregation level. In insurance based health care systems (as opposed to NHS-like systems) insurers will need certain patient data to control eligibility and reimburse either the patient or the health care provider. However, those data, although directly identifiable, can and should be fairly general.

Epidemiological researchers often need to know nearly all the data. Once entered into the research domain those data can be without direct identifiers but under a pseudonym. Before that stage, it is sometimes unavoidable that researchers or research assistants interact with doctors on the level

of patient records to assess eligibility for entering the research domain, to avoid doctor's bias when selecting patients, to purge records, etc.

26.
The privacy debate should be seen in a 'contextual context'. The use of patient data for quality assurance is not considered as alien to the individual function, the primary reason why those data were noted in the patient record. Yet the line between quality assurance and research is a blurry one. Without research the patient could not have been treated as he or she is at present. Researchers with patient data should not be seen as alien to the privacy context of patient care but as part of that context, if certain conditions are met.

27.
Health research with data differs fundamentally from interventional research which changes the justifiable expectation of the patient that he will be treated according to the best available standards. Any deviation from those standards should be thoroughly assessed on their acceptable risks and following that, based on informed consent of the patient concerned. With observational research with data, when the patient is still treated accorded to the best available standards, there is no risk for the patient involved. The patient does not even have to contribute actively. The research is carried out, as it were, at the background of the patient's treatment and usually after the treatment has ended.

28.
The results of this research will leave the research domain only as publications about patterns, leading to better treatment for future patients, or pointing at environmental, life style etc. risks, for the population. The main challenge to that assumption, insofar as relevant in this context, is twofold:

- Patient data are not safe in the research domain and researchers or third parties might still re-identity patients;
- Even if those data would be safe, publication about patterns jeopardise the 'group privacy' of parts of the population and might lead to anxiety with that groups, to discrimination of members that group by others, etc.

The first challenge has been referred to already. The second challenge would also apply if only anonymous data would be used in the research domain. It is a fundamental challenge to all epidemiological but also social research as such. About this the following, where the challenge itself is challenged, limited to health research.

29.
The challenge of group privacy:

- Presupposes that all people belong to one group, while social reality is far more complex.
- The most notorious case of patterns relating to groups is that of smokers versus non smokers. And hence the basic moral issue: opposite those who feel wronged or worried, are those whose lives might be saved by those results as they can take preventive measures themselves or the results are translated into health care for preventive health care, speeding up the diagnostic process etc. The latter weighs more.
- The same holds true if the outcomes concern certain inherent properties of persons

instead of practices which one can give up. There is first of all the translation of those outcomes in health care, resulting in offering treatment, or better disease management, preventive measures such as screening programs, etc.

- Sometimes general results of research will be translated outside health care in the legislation, statutes, by laws, policies etc. of governmental or private entities and then be applied to members of the group. Again smokers are a case in point. Such translation always involves various intermediary steps. In our European societies these are subject to the fabric of public and private law, including anti discrimination law, and can ultimately be challenged in Court. If the outcome would be that certain discrimination is not considered illegal and that outcome is unsatisfactory, the law should be challenged, not the results of research, if they are valid.

A civic society with a free press, where minorities are heard and can make themselves heard, where possible discriminatory practices, outside the very private realm, are subject to the challenges of public and private law, is the basic presumption. Additionally the European social system which its safety net of solidarity based social services and a solidarity based health care system which is immensely more than a simple safety net. In such a society invoking 'group privacy' as a barrier to research with data, is to some extent even anti-democratic and paternalistic to the possible 'groups' concerned. Again, this does not mean that anything goes. Not with respect to how data are handled nor with respect to how the research agenda is set. Only the former aspect can be discussed here.

30.
The conclusion is that patient data may be used for the 'public good', if certain conditions are met. Micromanagement by the patient resulting in informed consent on which data may be used for which research is not one of those conditions. The public good is seen neither as the aggregate of individual interests nor the public as whole, who might suffer as a whole. The public good or 'public interest' here is the protection of individuals against avoidable health threats to which they can not protect themselves. It is a public good to organise society in such a way that valuable individual interests are safeguarded when these interests are too vulnerable to protect themselves. The interests of future patients are as valuable as those of present patients. Yet, they are sufficiently more vulnerable here as, amongst many other things, they cannot negotiate with present patients about releasing data for better treatment in the future.
Hence, this is an argument which sees present patients as responsible citizens who profit as patients from the health care system as it is and may expected to contribute in their turn to better health for other patients, under the conditions mentioned earlier. A contrary conclusion, based on e.g. a conception of privacy as informational self-determination, would lead to the 'tragedy of anti-commons'.

31.
Yet, if this would lead to a barrier to health care as patients might nevertheless become afraid that their data will be misused, the contrary would be achieved of for what the present 'paradigm shift' is meant. Hence, patients should always be given the opportunity to opt out from using 'their' data for health research. That is only different if there is concrete statutory basis to submit data by health care providers.

32.
The concluding chapter looks back and discusses some of the remaining challenges. The focus in

the Report has shifted. From the conclusion that a TTP cannot solve all problems for good health research and that researchers will still need research exemptions to a proposal that they could need them less if, instead of assuring that data reach the research domain anonymously, it is assured that data are secure there, also against illegitimate re-identification.
Chapter 10 addressed more fundamental issues. It pleaded for a 'paradigm shift':

- The collective function of patient data in health care is not the exception to the individual function. Data are being processed or used for the collective function just as much as for the individual function. That collective function is equally important for patient care and the sustainability of equitable health care systems and improvement of health care as we know it;
- Researchers are not by definition 'strangers to the bedside'. In a contextual approach of privacy they are part of a continuum by which those closely involved with patient care try to improve this care;
- Patients may be expected to contribute to improvement of health care by allowing data collected for the individual function, to be used for health research, if certain conditions are met.

The first element explains the actual dataflow in all European health care systems much better than the present paradigm does. The following two elements explain the moral beliefs and professional dedication of (epidemiological) researchers, of many patient organisations and the practices in those EU countries which have used the research exemptions in Directive 95/46/EC to the max. They also express my view on a justifiable use of patient data. It can only be hoped that this will lead research in medical law in a new direction, and the discussion about the update of Directive 95/46/EC as well.

33.
The remainder of the concluding chapter discusses various practical aspects and 'what if' objections without going into detail. 'Necessity' and 'proportionality' of using data for research are essential elements of the proposed system as well. Yet, also here more trust in the professionalism of researchers can be warranted.
For the future it is submitted that the opt-out should also apply to the use of fully anonymous data. The difference between indirectly identifiable data and anonymous data has to a large extent become irrelevant.
This can only be implemented in the longer run, embedded in an obligation of health care providers to implement such an opt-out system and EHCR software which allows annotation of opt-out and following this opt-out through the course of the patient's treatment.

DUTCH PART

DE TTP ALS PANACEE?

1. INLEIDING

Wetenschappelijk onderzoek met patiëntgegevens is van essentieel belang voor de gezondheidszorg. Zo komt bijvoorbeeld kennis beschikbaar over aspecten van onze leefomgeving die onze gezondheid kunnen bedreigen of over de effecten van bestaande behandelingen.
De uitkomsten van zulk onderzoek komen beschikbaar in de vorm van publicaties waarin wel (patiënten)groepen worden onderscheiden maar individuen niet herkenbaar zijn. Deze uitkomsten worden – in de regel na nader, soms ook klinisch, onderzoek[1] en vaak ook maatschappelijke discussie – verwerkt in bijvoorbeeld de aanpassing van emissienormen of het aanpassen dan wel niet meer toepassen van bepaalde behandelingen. Voor degenen van wie deze gegevens worden verwerkt, komen de uitkomsten overigens meestal te laat. Anderen kunnen er op deze indirecte wijze wel van profiteren.
Een vroeg voorbeeld van zulk onderzoek is hoe in de 19e eeuw door John Snow tegen de toenmalige opvattingen in het verband werd ontdekt tussen bepaalde waterputten en de in bepaalde arme wijken van London endemisch voorkomende cholera.[2] Een meer recent voorbeeld is het ontdekken van het verband tussen roken en kanker en bepaalde hart- en vaatziekten. Er zijn tal van andere voorbeelden, zowel van onderzoek naar de gevolgen van omgevings- of gedragsvariabelen en het al dan niet optreden van ziekten als van onderzoek naar behandelingen en hun latere effecten als iemand eenmaal ziek is geworden.[3]

Zulk onderzoek met patiëntgegevens dient te voldoen aan de relevante wetgeving, voor Nederland met name de Wet op de geneeskundige behandelingsovereenkomst (WGBO) en de Wet bescherming persoonsgegevens (Wbp). Dat geldt overigens niet alleen voor onderzoek met patiëntgegevens maar voor onderzoek met persoonsgegevens in het algemeen, zij het dat dan de WGBO niet van toepassing is en de uit de Wbp volgende voorwaarden in een aantal situaties minder stringent zijn.[4]

De Gedragscode Gezondheidsonderzoek[5] van de FEDERA[6] werkt de uit de WGBO en Wbp volgende voorwaarden nader uit. Inmiddels is de verklaring van overeenstemming met de geldende wetgeving van het College bescherming persoonsgegevens (CBP) ex artikel 24 Wbp geëxpireerd. Het CBP heeft verzocht om in een volgende versie aandacht te besteden aan een aantal onderwerpen die in de huidige Gedragscode niet aan de orde komen, zoals de inschakeling van een TTP bij gezondheidsonderzoek.

Alvorens de TTP (al dan niet) in de nieuwe Gedragscode te verwerken, wenste de FEDERA van MedLawconsult een analyse van de mogelijke functie van een TTP bij zorgonderzoek. Aan dat verzoek beoogt deze notitie te voldoen.
De notitie is als volgt opgebouwd:
Eerst komt een initiële verkenning waarin de vraagstelling met betrekking tot een TTP bij zorgonderzoek wordt gespecificeerd. Dan volgt een analyse van de voorwaarden voor wetenschappelijk onderzoek met patiëntgegevens. Immers, de TTP zou een rol moeten hebben in het kader van die voorwaarden. Uit deze twee hoofdstukken blijkt dat de status van 'gecodeerde gegevens' in de discussie een sleutelrol vervult. Daarop wordt vervolgens in een langer hoofdstuk ingegaan. Pas nadat die initiële verkenningen zijn uitgevoerd, kan worden ingezoomd op de rol van een TTP bij zorgonderzoek, de voor- en nadelen daarvan en de vraag wanneer deze wel en wanneer deze niet ingeschakeld zou hoeven te worden.

Deze hoofdstukken sluiten zoveel mogelijk aan bij het geldende recht. In het Engelse deel van dit rapport worden daarbij enkele vraagtekens geplaatst en worden oplossingsrichtingen aangedragen. Ik schets een toekomstvisie die het beperkte kader van de 'TTP, wat kunnen wij er mee', overstijgt.

De FEDERA is bekend met mijn kritische benadering van het geldende recht[7] en vroeg mij waar-

schijnlijk niet voor niets. Voor de lezer zal het de vraag zijn in hoeverre dit standpunt de beschrijving van het geldende recht heeft beïnvloed. Dat is een terechte vraag. Recht is een normatieve activiteit en ook de beschrijving (dus nog los van de beoordeling) van het geldende recht kan niet geheel los worden gezien van normatieve uitgangspunten. Dat blijkt dan met name uit het leggen van bepaalde accenten. Zo wordt relatief veel aandacht besteed aan de ervaringen van onderzoekers. Maar ook als (juridische) feiten geen feiten zijn, moet je wel proberen zo open en eerlijk mogelijk te zijn. Alles proberen op te merken en te benoemen, niet voorbijgaan aan bepaalde 'feiten' die minder passen[8] en verklaren waarom je het zo benoemt. Daarnaast, maar dat is accessoir, dient het betoog consistent en logisch te zijn.

2. DE TTP, EEN EERSTE VERKENNING

Het begrip 'trusted third party' (TTP) deed eind vorige eeuw zijn intrede in de financiële wereld.[9] Om verschillende redenen bestond behoefte aan een onafhankelijke derde die fungeerde als tussenpersoon bij met name bancaire transacties om de identiteit van de partijen te verifiëren en te garanderen waartussen de transactie plaats vond. In de financiële wereld is echter nooit eenduidigheid ontstaan over de voorwaarden waaronder een natuurlijke- of rechtspersoon een 'trusted third party' kan worden genoemd. Die voorwaarden zijn deels contextafhankelijk.

Ten behoeve van de uitvoering van de structuur- en financieringswetgeving[10] in gezondheidszorg wordt de TTP de laatste jaren steeds meer ingeschakeld. Het begrip heeft daarbij een geheel andere betekenis gekregen. Kort gezegd, niet meer om de identiteit van een betrokkene[11] te verifiëren, maar om deze te *maskeren*. Een partij als bijvoorbeeld het College voor zorgverzekeringen (CVZ) of het Ministerie van VWS heeft unieke gegevens over verzekerden of 'zorgconsumenten' nodig zonder over persoonsgegevens te mogen beschikken. De TTP waarborgt dat persoonsgegevens van de verstrekkende partij (zorgverzekeraars of zorginstellingen) worden ontdaan van identificerende kenmerken en de ontvangende partij wel inhoudelijke gegevens krijgt over unieke personen (soms cumulatief, over dezelfde personen dus) maar dat deze gegevens niet kunnen worden herleid tot de betrokkenen (zie ook sectie 3.2).

Ook bij het gezondheidsonderzoek wordt de TTP in toenemende mate genoemd. Het CBP stelde enkele jaren geleden in diens rapport over zorgregistraties dat het mogelijk zou moeten zijn om de identiteit van de betrokkenen in die registraties beter te maskeren zonder aan de onderzoekswaarde van de registraties afbreuk te doen.[12] Eerder verscheen een rapport van de voorganger van het CBP, de Registratiekamer, over zogenaamde privacy enhancing technologies (PET's) met als subtitel 'The path to anonimity'.[13] De TTP zou bij gezondheidsonderzoek als een meest verstrekkende vorm van een PET kunnen worden gezien. Ook bij de recente 'invitational conference' van de Belgische Commissie voor de Bescherming van de Persoonlijke Levenssfeer over privacy en wetenschappelijk onderzoek werd eveneens sterk ingezet op ICT als de panacee voor het opheffen van de spanning.[14]
Zoals opgemerkt heeft het CBP in de discussie met de FEDERA over de verlenging van de geldigheid van de Gedragscode Gezondheidsonderzoek de TTP als een te behandelen onderwerp genoemd, terwijl de recente brief van de Minister van VWS aan de beide Kamers stelt dat om het koppelen van buiten het CBS gegenereerde databestanden mogelijk te maken, een TTP nodig is om aan de voorwaarden van het CBP te voldoen.[15]

De mogelijke functie van de TTP voor zorgonderzoek is echter nog niet systematisch beschreven. Dat blijkt uit het nagenoeg ontbreken van literatuur over de functie en voorwaarden van de TTP bij gezondheidsonderzoek of in de zorg in het algemeen. Een afweging van de voor- en nadelen en bepaling van het 'indicatiegebied' van de TTP bij gezondheidsonderzoek heeft dan ook evenmin plaatsgevonden.
Deze notitie beoogt dat te veranderen en duidelijkheid te scheppen over:
- De eigenschappen waaraan een 'echte' TTP in het algemeen moet voldoen.
- De (mogelijke) functie van een TTP voor gezondheidsonderzoek.
- Wat de mogelijke nadelen van een TTP zijn.
- Wanneer het wel en niet zinvol is om een TTP bij gezondheidsonderzoek in te schakelen.
- Welke alternatieven voor een 'echte' TTP beschikbaar zijn als de nadelen groter blijken dan de voordelen.

Deze onderwerpen komen later in deze notitie aan de orde. Eerst worden enkele tussenstappen gemaakt. De TTP voor gezondheidsonderzoek moet uiteraard worden gezien in het licht van de voorwaarden om dergelijk onderzoek te kunnen verrichten en aan de voorwaarden waaraan gegevens moeten voldoen om bruikbaar te zijn voor wetenschappelijk onderzoek. Daarover eerst een

kort hoofdstuk. Aansluitend daarop een langer hoofdstuk over de status van gecodeerde gegevens. Zoals zal blijken betekent de omwerking van persoonsgegevens tot anonieme gegevens via een TTP namelijk dat deze anonieme gegevens ook gecodeerde gegevens zijn.

3. DE VOORWAARDEN VOOR GEZONDHEIDSONDERZOEK

Voor het verwerken van patiëntgegevens bij gezondheidsonderzoek door een ander dan de behandelend hulpverlener is in beginsel de uitdrukkelijke toestemming van de betrokkene nodig. De uitzonderingen op dit beginsel zijn in de Gedragscode Gezondheidsonderzoek[16] geformuleerd. Deze uitzonderingen zijn gebaseerd op de artikel 7:458 BW en artikel 23, tweede lid, Wbp.
Kort samenvattend komen deze uitzonderingen inhoudelijk op het volgende neer:

a. Het gebruik van anonieme gegevens mag (momenteel) altijd. Wel beschouwd is dit geen uitzondering maar een (tweede) uitgangspunt. Men mag dan uiteraard niet bestanden met anonieme gegevens zodanig koppelen dat toch indirect of zelfs direct identificerende gegevens worden verkregen.
b. Onder voorwaarden is het gebruik van (al dan niet gecodeerde) gegevens mogelijk die niet anoniem zijn maar waarvan herleiding redelijkerwijs wordt voorkomen. Dit zijn geen direct identificerende gegevens, dus gegevens zonder een naam of met andere identificatoren waarmee de betrokkene direct kan worden herkend.
   De primaire voorwaarde om zulke gegevens te mogen gebruiken is dat het vragen van toestemming in redelijkheid niet van de onderzoekers kan worden verlangd. De facto betekent dit dat achteraf aan een grote groep betrokkenen om toestemming zou moeten worden gevraagd. Daarnaast gelden additionele voorwaarden, zie de tekst hierna.
c. Onder strikte voorwaarden is ook het gebruik van direct identificerende gegevens mogelijk, tot en met zelfs inzage in het patiëntendossier. De primaire voorwaarde is dat het vragen van toestemming onder de gegeven omstandigheden niet mogelijk is. De Gedragscode Gezondheidsonderzoek werkt deze omstandigheden verder uit.

Zowel bij b en c gelden als additionele voorwaarden dat:

- Het onderzoek een algemeen belang moet dienen.
- Zonder het gebruik van deze gegevens het onderzoek niet kan worden uitgevoerd.
- Het onderzoek is beschreven in een onderzoeksprotocol.
- Een METC dit onderzoeksprotocol tevoren heeft beoordeeld en een advies heeft uitgebracht over de zorgvuldigheid ervan.[17] [18]
- De betrokkene tegen zulk gebruik van diens gegevens tevoren geen bezwaar heeft gemaakt. Een algemeen 'geen bezwaar'-systeem aan de 'poort' van de zorginstelling (of zorgverzekeraar) voldoet hiertoe.

Deze voorwaarden worden door onderzoekers als betrekkelijk restrictief ervaren. Met name worden genoemd:

1. Het tijdens de behandeling vragen van uitdrukkelijke toestemming voor later gebruik van de patiëntgegevens voor wetenschappelijk onderzoek komt er niet van. Patiënten en behandelaars lijken dan wel wat anders te bespreken te hebben.
2. Het vragen van toestemming (vooraf of achteraf) leidt tot 'bias' bij het onderzoek.
3. Het geen bezwaar-systeem aan de poort wordt door zorginstellingen niet goed ingevuld.
4. De voorwaarden om van anonieme gegevens te kunnen spreken zijn bijzonder zwaar en worden door veel onderzoekers ook als beledigend voor hun integriteit ervaren.

Ad 1
Een berucht voorbeeld is het einde van de in Duitsland bestaande kankerregistratie toen daar uitdrukkelijke toestemming voor het verstrekken van gegevens aan deze registraties werd geïntroduceerd.[19] Ook thans blijkt dat artsen vaak wel bereid zijn mee te werken aan onderzoek met gegevens uit hun praktijken zolang de belasting maar niet te groot is. Het vragen van toestemming aan elke patiënt wordt als een belasting ervaren met name omdat de dialoog van patiënt en arts op iets anders is gericht dan secondair gebruik van gegevens, namelijk voor de patiënt: wat is mijn diagnose en wat zijn de behandelopties.[20]

Ad 2
Met bias wordt gedoeld op een vertekening in de uitkomsten van het observationeel wetenschappelijk onderzoek doordat slechts een bepaalde groep personen toestemming geeft om hun gegevens te mogen gebruiken. Er is ruim voldoende empirisch onderzoek dat zulke bias inderdaad kan optreden, met name ten nadele van zwakkere sociaal-economische groepen, indien achteraf toestemming wordt gevraagd.[21] Onderzoekers willen deze bias vermijden, teneinde tot valide resultaten te komen die op alle patiënten van toepassing zijn. De aanmerkelijke kans op bias geldt echter niet als een reden waarom 'vragen van toestemming in redelijkheid niet kan worden verlangd' of 'niet mogelijk is'.

Ad 3
Een geen bezwaar-systeem veronderstelt voldoende duidelijke vermelding in de algemene informatiefolder of algemene voorwaarden en een laagdrempelige mogelijkheid voor nadere inlichtingen of om bezwaar te maken. Voorts dient het eventuele bezwaar in de ICT-systemen te kunnen worden verwerkt.
Voor veel zorginstellingen en zorgverzekeraars was gezondheidsonderzoek echter een nogal abstracte aangelegenheid waar men geen direct belang bij dacht te hebben. Niemand leek 'probleemeigenaar' van de vormgeving van het geen bezwaar-systeem. Dat kwam er dan ook niet. Dat blijkt ook uit het feit dat NEN-norm 7510 niet behandelt dat het ICT-systeem ook in zo'n bezwaar tegen 'nader gebruik' van de gegevens (of lichaamsmateriaal) voor wetenschappelijk onderzoek moet kunnen voorzien.[22] Onderzoekers zitten aan het einde van de informatieketen en bezitten niet de mogelijkheden om dit aan de poort vorm te geven.[23]
Overigens lijkt dit momenteel iets ten goede te veranderen. Zorginstellingen en zorgverzekeraars zijn zich meer en meer bewust van de mogelijkheden van gezondheidsonderzoek om de kwaliteit van zorg te kunnen verbeteren.

Ad 4
Zoals in het volgende hoofdstuk nader zal worden behandeld, wordt de drempel om van anonieme gegevens te kunnen spreken erg hoog gelegd. Daar worden ook voorbeelden genoemd van situaties waarbij onderzoekers in gemoede meenden met anonieme gegevens te werken, maar door toezichthouders werd gesteld dat hiervan geen sprake was vanwege de mogelijkheid voor de betrokken onderzoekers om op oneigenlijke wijze alsnog de identiteit van de patiënt te achterhalen.

Een en ander zou geen probleem hoeven te zijn indien onderzoekers inderdaad met volledig geanonimiseerde gegevens zouden kunnen werken. Maar dat is niet het geval. Ik kom daar nog op terug. Wel is het zo dat onderzoekers niet zijn geïnteresseerd in de personen achter de gegevens, maar in de patronen van, kort gezegd, eventuele blootstelling, behandeling en uitkomsten in termen van (on)gezondheid die via die gegevens zichtbaar worden. Hierna kortweg het adagium: geen personen maar patronen.
Daartoe is tweeërlei nodig:
- onderzoeksgegevens die voldoende en de juiste informatie bieden om de onderzoeksvraag te kunnen beantwoorden;
- een uniek kenmerk voor elk individu waardoor op verschillende individuen betrekking hebbende onderzoeksgegevens van elkaar kunnen worden onderscheiden.

Die onderzoeksgegevens zijn vaak onvolledig of onduidelijk. Dan moet men op een of andere manier terug naar de bron om nadere informatie over die betrokkene op te vragen. Of het onderzoek bestaat uit combineren van gegevens uit verschillende bronnen. Ook als de identiteit van de betrokkene daarbij wordt gemaskeerd, is het wel noodzakelijk om de gegevens over een individu uit de ene bron te koppelen aan gegevens over dezelfde individu uit de andere bron. De wijze van aanmaken van deze pseudo-identiteit moet dan door beide bronnen worden gedeeld. Vaak gaat het daarbij mis. In veel van die gecombineerde onderzoeksdatabases komen dubbelingen of 'meerlingen' voor, die vervolgens handmatig moeten worden uitgezuiverd.[24]

Gebruik van een TTP zou ertoe kunnen leiden dat gegevens die direct of indirect identificerend zijn, als anoniem mogen worden beschouwd. Dat zou een groter en gemakkelijker gebruik van gegevens mogelijk maken. Geen personen maar patronen tenslotte. Indien de TTP zulke technieken in huis heeft dat de gegevens inderdaad volgens de – hierna te behandelen strikte normen – de onderzoeker anoniem bereiken, terwijl wel nadere gegevens kunnen worden opgevraagd of dat inderdaad unieke personen kunnen worden gekoppeld, zou voor bepaald onderzoek veel zijn bereikt.

Naarmate meer en/of grotere bestanden met elkaar worden gekoppeld (bijvoorbeeld bij 'datasharing')[25] zal dit meer gaan spelen, zeker als beperkingen in de toestemming dit bredere delen met vaak een net andere vraagstelling moeilijk maken.

Inherent aan zulke koppeling is dat deze gegevens zijn gecodeerd. In plaats van de direct of indirect identificerende gegevens[26] is er een uniek door de TTP gegeneerd pseudoniem waaraan de onderzoeksgegevens als het ware zijn opgehangen. Dat proces is herhaalbaar als nieuwe gegevens moeten worden opgevraagd. Met name door een deel van de juristen in het Nederlandse gezondheidsrecht wordt evenwel gesteld dat gecodeerde gegevens per definitie geen anonieme gegevens kunnen zijn. Op die kwestie zal in het volgende hoofdstuk worden ingegaan.

4. PERSOONSGEGEVENS VERSUS ANONIEME GEGEVENS

4.1 Het begrip persoonsgegevens in de WGBO en de Wbp

De WGBO stelt voorwaarden aan het gebruik van (niet anonieme) patiëntgegevens voor wetenschappelijk onderzoek, de Wbp aan het gebruik van bijzondere persoonsgegevens. Beide begrippen moeten uiteraard hetzelfde worden uitgelegd.[27] Niet anonieme patiëntgegevens in de zin van de WGBO zijn persoonsgegevens in de Wbp. Geen persoonsgegevens in de zin van de Wbp, zijn anonieme patiëntgegevens in de zin van de WGBO. Elke andere conclusie is absurd.

Waar artikel 7:458 BW (WGBO) de in het vorige hoofdstuk onder b besproken versoepeling inhoudt voor onderzoek met patiëntgegevens, gaat het dus nog steeds om persoonsgegevens in de zin van de Wbp. Zo staat het er ook. Het betreft gegevens waarvan 'herleiding redelijkerwijs wordt voorkomen'. Niet het Wbp-begrip (zie hierna) voor anonieme gegevens: "herleiding is zonder onevenredige tijd en moeite niet mogelijk". Zou het bij 7:478 BW lid 1 onder b om anonieme gegevens gaan, was de bepaling trouwens ook overbodig.[28] Hetzelfde geldt voor artikel 23 tweede lid Wbp die met artikel 7:458 BW overeenkomende versoepeling voor gebruik van gegevens voor wetenschappelijk onderzoek zonder uitdrukkelijke toestemming mogelijk maakt. Het betreft dan uiteraard onderzoek met persoonsgegevens.

4.2 Anonieme gegevens in het algemeen

Van anonieme gegevens is slechts sprake als herleiding redelijkerwijs niet mogelijk is. In de discussie over Richtlijn 95/46/EC waarop de Wbp is gebaseerd, wordt gesteld dat om dit te bepalen rekening moet worden gehouden met: "all the means likely reasonable to be used either by the controller or by any other person to identify the said person."[29]
Dat 'likely reasonably to be used' lijkt een concrete toetsing in te houden die rekening houdt met de omstandigheden waaronder de gegevens worden verwerkt. De facto gaan toezichthouders echter veelal van een abstracte benadering uit. Als herleiding praktisch op enigerlei wijze denkbaar is, dan zijn de gegevens niet anoniem.
Daarbij is het van belang een onderscheid te maken tussen:
- indirecte herleiding in verband met het aggregatieniveau (a);
- indirecte herleiding op een andere manier, met name de codering van de gegevens (b).

Ad a
Zo worden vier cijfers van de postcode, een zeldzame ziekte, geslacht en bijvoorbeeld vijfjaarsklassen al snel als indirect herleidbaar beschouwd. Toch is het lastig voor te stellen hoe iemand dan achter de naam van de betrokkene zou kunnen komen zonder heus recherchewerk te verrichten. Een onderzoeker zal daartoe niet overgaan. Dat is echter wel het uitgangspunt van de interpretatie van het begrip indirect herleidbaar door toezichthouders,[30] die het begrip als het ware niet contextafhankelijk invullen. Met niet contextafhankelijk wordt gedoeld op het uitgangspunt dat niet relevant is wie over de gegevens beschikt, bijvoorbeeld een marketingbureau dat eventueel zou kunnen verdienen aan het persoonlijk benaderen van de betrokkenen of onderzoekers die uitsluitend zijn geïnteresseerd in patronen en voorts aan beroepscodes en interne reglementen zijn gebonden.

Ad b
Dit speelt vooral bij op een of andere manier gecodeerde gegevens. Twee voorbeelden uit de praktijk ter illustratie:[31]

> In een onderzoeksdatabase zijn patiënten opgenomen onder patiëntnummer van de aanleverende praktijken alsmede onder het nummer van die praktijk. Slechts twee senior onderzoekers hebben toegang tot deze database. Het CBP concludeerde dat de database geen anonieme gegevens bevat. De desbetreffende onderzoekers zouden de praktijken kunnen bellen om achter een naam te komen. Dat zoiets in strijd met de interne regels zou zijn en het einde van hun carrière zou betekenen als het uit zou komen, dat er geen

redelijk belang bij is (wat heb je er aan?) en dat ook vanuit de aanleverende praktijk dan de geheimhoudingsplicht zou worden geschonden, speelde voor het CBP geen rol.

In een database zijn patiënten met de inhoudelijke gegevens via een éénwegcodering opgenomen. Een zeer beperkt aantal medewerkers van de organisatie heeft toegang tot de database. Zij hebben ook de toegang tot de codering, dus van NAW-gegevens naar het codenummer. Andersom is de codering goed beveiligd, er wordt ook regelmatig een audit op uitgevoerd. Het is dus ook voor deze medewerkers niet mogelijk om van het codenummer naar de patiënt terug te gaan. Zij zouden echter wel NAW-gegevens kunnen invoeren (van de buurman bijvoorbeeld) en dan kunnen zien of deze in de database voorkomt. Om die reden concludeerde het CBP dat de gegevens op dat niveau niet als anoniem kunnen worden aangemerkt.

In beide gevallen werden geheimhoudingsverklaringen en interne regels voor gebruik van de database, die het door het CBP mogelijk geachte gebruik uiteraard niet toestaan, als onvoldoende aangemerkt. Met name de opvatting van het CBP in de eerste casus heeft bij onderzoekers tot verontwaardiging geleid. Men vond het een belediging voor hun integriteit.[32]

4.3 Gecodeerde gegevens

Over de vraag of gecodeerde gegevens als zodanig als anonieme gegevens mogen worden aangemerkt, bestaat in Nederland al geruime tijd discussie. Ploem stelt dat dit steeds persoonsgegevens zijn, waarvoor dan wel een lichter regiem mag gelden voor gebruik bij wetenschappelijk onderzoek.[33] De heersende opinie is echter dat deze gegevens geen persoonsgegevens zijn, althans niet per definitie.[34]

De Opinie over 'The concept of personal data' (4/2007) van de zogenaamde 'Article 29 Data Protection Working Party'[35] heeft naar mijn mening aan de discussie een einde gemaakt. Gecodeerde gegevens *kunnen* wel degelijk anonieme gegevens zijn. De cursivering betekent wel dat een genuanceerde benadering is vereist.
Het is onjuist om gecodeerde gegevens per definitie bij de ene of de andere categorie, dus wel persoonsgegevens of geen persoonsgegevens, in te delen. Of gecodeerde gegevens anonieme gegevens opleveren hangt af van de twee eerdergenoemde factoren:
- de veiligheid van het coderingsmechanisme;
- het aggregatieniveau van de gegevens die onder het codenummer zijn opgenomen. Met andere woorden: zijn de gegevens al dan niet op andere wijze indirect herleidbaar.

Over het coderingsmechanisme het volgende. In 2009 heeft het CBP een onderzoek bij de administratie van het College voor zorgverzekeringen (CVZ) uitgevoerd om na te gaan of de gegevens die in het kader van de AWBZ-brede zorgregistratie (AZR)[36] worden verwerkt in overeenstemming zijn met de Wbp en aanverwante wet- en regelgeving (CBP 2009a). In het kader van dit onderzoek formuleerde het CBP een aantal voorwaarden, waaronder bij toepassing van codering (pseudonimisering) geen sprake is van het verwerken van persoonsgegevens:
1. Er wordt (vakkundig) gebruik gemaakt van pseudonimisering, waarbij de eerste encryptie plaatsvindt bij de aanbieder van de gegevens.
2. Er zijn technische en organisatorische maatregelen genomen om herleidbaarheid van de versleuteling ('replay back') te voorkómen.
3. De verwerkte gegevens zijn niet indirect identificerend.
4. In een onafhankelijk deskundig oordeel (audit) wordt vooraf en daarna periodiek vastgesteld dat aan de voorwaarden 1, 2 en 3 is voldaan.
5. De pseudonimiseringsoplossing dient helder en volledig te zijn beschreven in een openbaar document, zodat een betrokkene kan nagaan welke garanties de gekozen oplossing biedt.

Met name aan de vierde voorwaarde zal bij veel gezondheidsonderzoek niet snel zijn voldaan. De vraag is of dat ook steeds nodig is. Daarop wordt in hoofdstuk 8 nog teruggekomen.

4.4 Tweeweg gecodeerde gegevens

In de Nederlandse discussie werd meestal[37] geen onderscheid gemaakt tussen éénweg- en tweewegcodering. Bij éénweg is er slechts één sleutel, van de NAW-gegevens naar het codenummer. Bij tweeweg is er nog een sleutel, namelijk van het codenummer terug naar de NAW-gegevens en dus de identiteit van de betrokkene. Kan bij tweeweg gecodeerde gegevens ook nog van anonieme gegevens worden gesproken?
De genoemde Opinie van de Article 29 Working Party biedt over deze vraag eveneens duidelijkheid. Het feit dat gegevens ergens in de keten kunnen worden herleid, betekent niet dat zij niet elders in de keten als anonieme gegevens kunnen worden aangemerkt. In theorie kunnen tweeweg gecodeerde gegevens daarom als anonieme gegevens worden aangemerkt. De voorwaarden voor de versleuteling zijn dezelfde als genoemd in sectie 4.3.
De Opinie merkt echter op: "the risks of external hack, the likelihood that someone within the sender's organization – despite his professional secrecy – would provide the key (to the receiver, *EBvV*) and the feasibility of indirect identification are factors to be taken into account …".[38] Met andere woorden, de eerder genoemde abstracte kans op herleiding geldt ook hier. En omdat die kans groter is, vanwege het feit dat iemand ergens in het proces toegang heeft tot de tweede sleutel (de sleutel terug van codering naar identiteit), zullen de veiligheidsprocedures dus nog strenger moeten zijn.

De Opinie lijkt hierin echter niet altijd even consequent te zijn. Er wordt een casus besproken van een clinical trial waarbij de zogenaamde 'case record forms' (CRF)[39] zoals dat hoort gecodeerd naar de externe sponsor worden gezonden. De Opinie concludeert dat deze CRF's bij de sponsor als anonieme gegevens mogen worden aangemerkt. Deze CRF's zijn tweeweg gecodeerd. De onderzoeker moet naar de patiëntgegevens terug kunnen bijvoorbeeld als de sponsor op basis van de geaggregeerde gegevens een bepaalde bijwerking rapporteert. Al helemaal indien een onderzoeker een ernstig ongewenst voorval met een proefpersoon meldt en deze inderdaad aan de trial medicatie[40] blijkt gerelateerd. Dan moeten de andere deelnemende centra dat voor hun patiënten zo spoedig mogelijk weten.[41] Dat gebeurt via de sponsor. De codering is echter in het algemeen nogal 'primitief'.[42] Bijhalve bij zeer grote trials is het niet meer dan een lijst bij de behandelend onderzoekers[43] met CRF-nummers en een corresponderende – uiteraard uitsluitend hen bekende – lijst met patiënten. De door de industrie ingeschakelde onafhankelijke monitor heeft bovendien het recht (en de plicht) om steekproefsgewijs aan de hand van de CRF's het dossier van de proefpersoon in te zien om te controleren of de CRF wel correct is ingevuld.[44] Hier kan in theorie eveneens het risico van een 'internal hack' bestaan of het vervagen van de grens tussen monitor en sponsor. Van privacyschendingen doordat de industriële sponsor op deze wijze de identiteit van de patiënt te weten zou zijn gekomen, is overigens nog nooit gebleken.

Het is niet geheel duidelijk waarom de Working Party hier tot een andere conclusie over anonimisering komt dan uit de eerder genoemde strikte voorwaarden lijkt te volgen. Mogelijk heeft men overwogen dat de industrie niet de identiteit van de proefpersoon *mag* achterhalen. Dat volgt onder meer uit de researchovereenkomst. En voorts dat de behandelend arts-onderzoeker of de betrokken verpleegkundigen deze identiteit niet mogen verstrekken. Dat volgt uit hun beroepsgeheim. In dat geval zou de Working Party echter een concrete benadering hebben gevolgd. Dat wil zeggen niet uitgaande van abstracte risico's, maar uitgaande van de concrete risico's gezien de normatieve inbedding van zender en ontvanger. Een dergelijk standpunt zou consequenties hebben voor de eerder behandelde casus. En voor veel meer van dergelijke gegevensstromen waarbij de privacy aan de bron goed wordt beschermd en onderzoekers de onderzoeksgegevens onder een naar hun beste weten 'anoniem' pseudoniem verkrijgen, zonder dat daar complexe systemen voor zijn ontwikkeld.
Het kan ook zijn dat de Working Party zich bij het voorbeeld onvoldoende bewust was van de praktijk en meer complexe ICT-systemen voor ogen had.

In het Engelse deel wordt betoogd dat een meer concrete benadering aanvaardbaar zou moeten

zijn. Op dit moment is voldoende de conclusie dat naar geldend recht ook tweeweg gecodeerde gegevens bij de ontvanger als anonieme gegevens kunnen worden aangemerkt, in ieder geval als bij die codering stevige beveiligingsprocedures zijn ingezet.

4.5 Vertaling naar de FEDERA Gedragscodes

4.5.1 De Gedragscode Gezondheidsonderzoek

De Gedragscode Gezondheidsonderzoek maakt in navolging van de WGBO (en de Wbp) onderscheid tussen:
- anonieme gegevens;
- gegevens waarvan herleiding redelijkerwijs wordt voorkómen;
- direct herleidbare gegevens.

De vraag is met name hoe nu de tweede categorie moet worden gekarakteriseerd. Hierover bestaat ook in de literatuur verwarring.[45] Het zijn in ieder geval geen anonieme gegevens. Anders was de tussencategorie overbodig (zie ook sectie 4.1). Omdat in de discussie naar de aanloop van artikel 7:458 BW wel is gesproken van codering voor deze categorie van gegevens (zie hiervoor), is er gemakshalve van uitgegaan dat het hier om gecodeerde gegevens moet gaan *en* dat gecodeerde gegevens dus geen anonieme gegevens zijn.
Beide aannames zijn onjuist. Zoals de Gedragscode al zegt gaat het om 'al dan niet gecodeerde gegevens'. Om in de tussencategorie terecht te komen moet een aantal privacyslagen met deze persoonsgegevens worden gemaakt. Die kunnen codering inhouden maar dat hoeft niet. De NAW-gegevens kunnen bijvoorbeeld ook gewoon worden verwijderd, zonder te worden vervangen door een codenummer. Het aggregatieniveau is dan echter nog zodanig dat indirect identificerende gegevens overblijven. Als codering wordt toegepast, wordt niet voldaan aan de genoemde voorwaarden om van gecodeerde anonieme persoonsgegevens te spreken. Bijvoorbeeld het coderingsmechanisme is niet optimaal beveiligd (zie ook het voorbeeld in par. 4.2) of, zelfs als die codering wel optimaal is, het aggregatieniveau is nog steeds te specifiek.
Maar als wel aan de genoemde voorwaarden is voldaan, is sprake van anonieme gegevens en vallen deze gegevens dus niet meer in de tussencategorie, ook al zijn ze gecodeerd.
Dit lijkt mij eigenlijk vrij elementaire logica en ik zie niet in waarom de verwarring nog verder zou moeten voortbestaan.

4.5.2 In de Code Goed Gebruik (FEDERA 2011)

In deze Gedragscode wordt een geen bezwaar-systeem voorgesteld voor 'nader gebruik' ten behoeve van wetenschappelijk onderzoek met gecodeerd anoniem lichaamsmateriaal. Een van de bezwaren tegen de eerdere Gedragscode uit 2002 (FEDERA 2002) was vanuit gezondheidsrechtelijke kring dat nu voor onderzoek met lichaamsmateriaal een lichter regiem zou gelden dan voor onderzoek met gegevens (Olsthoorn 2003).
Zoals uit het bovenstaande blijkt, is dat bezwaar onjuist. Gecodeerd anoniem is anoniem, zowel voor de gegevens als voor het lichaamsmateriaal. Er geldt dus voor laatstgenoemde categorie geen lichter regiem, maar juist een zwaarder. Anonieme gegevens mogen naar huidig recht altijd worden gebruikt, ook indien gecodeerd. Anoniem lichaamsmateriaal (ook indien niet gecodeerd) evenwel slechts op basis van geen bezwaar.
Een andere kwestie is dat in de Code Goed Gebruik van 2002 de voorwaarden voor gecodeerd-anoniem niet werden geëxpliciteerd. Veel van hetgeen onderzoekers gecodeerd anoniem noemen, zal niet voldoen aan de voorwaarden die eerder zijn beschreven.

Overigens zijn er tussen anonieme gegevens en (gecodeerd) anoniem lichaamsmateriaal wel twee andere verschillen die voor de regulering van belang kunnen zijn.
Het eerste verschil is dat aan lichaamsmateriaal nieuwe gegevens kunnen worden ontleend. Via die gegevens zou men de reeds bestaande gegevens kunnen verrijken en dus alsnog tot herleiding kunnen komen. Onder specifieke omstandigheden is het zelfs mogelijk om in een pool van DNA-

gegevens personen steeds tekunnen onderscheiden (Homer 2008). Deze kwestie wordt uitvoerig besproken in bijlage 1 paragraaf 2.3 van de nieuwe versie van de Code Goed Gebruik (FEDERA 2011) en Riegman 2011. De conclusie is dat het onderscheiden welk DNA bij welk individu behoort niet zonder meer betekent dat die individu daarmee ook wordt geïdentificeerd in de zin van de Wbp (of de WGBO). Voor dat laatste zou veel meer moeten gebeuren en dat dit zou gebeuren is naar mijn mening terecht getypeerd als 'extremist' (Lowrance 2007, zie ook Riegman 2011).

Het tweede verschil is het volgende. Bij anonieme gegevens doet het er niet toe of dat via een éénweg- of tweewegcodering gebeurt, zolang de codering maar veilig blijft. Bij gecodeerd anoniem lichaamsmateriaal maakt éénweg of tweeweg gecodeerd wel verschil. Zoals opgemerkt leidt de analyse van het lichaamsmateriaal tot nieuwe gegevens. Dat kunnen 'bevindingen' zijn over bepaalde risicofactoren bij de donor. Bij éénwegcodering is het niet mogelijk om deze 'bevindingen' terug te koppelen naar de verstrekker van het materiaal. Immers, die weg is uitgesloten. Ook de verstrekker kan aan de hand van het codenummer niet meer de identiteit van de donor achterhalen. Bij tweewegcodering is dat laatste in beginsel wel mogelijk. Afhankelijk van hoe men aankijkt tegen de status van zulke 'bevindingen', de eventuele bezwaarlijkheid van deze en de wenselijkheid dat donoren zeggenschap hebben over individuele terugkoppeling, zal dat gevolgen hebben voor de regulering en het toestemmingssysteem voor het gebruik van lichaamsmateriaal voor wetenschappelijk onderzoek. Op die kwestie zal hier nu niet worden ingegaan. Deze notitie betreft de status van gegevens.

4.6 'Spontane herkenning'

Naast een onvoldoende hoog aggregatieniveau en onvoldoende beveiligde codering, wordt soms nog een derde mogelijkheid genoemd voor ongewenste herleiding, namelijk 'spontane herkenning'. Daarmee wordt gedoeld op het feit dat een bepaalde persoon uitsluitend via bepaalde gegevens die in het algemeen anoniem zouden zijn, de betrokkene herkent omdat de persoon juist die gegevens al van de betrokkene weet en dus met de betrokkene kan verbinden. Een voorbeeld is hoe op een college een bepaalde casus werd behandeld en een van de studenten daarin een vriendin herkende.
Nu is spontane herkenning uitsluitend dan een probleem indien men daarbij, naast de gegevens waaraan men de betrokkene herkent, meer gegevens verneemt die men niet had behoren te vernemen. Bijvoorbeeld op dat college herkent de student diens vriendin aan bepaalde contextuele omstandigheden en hoort daarbij over een aandoening die hij niet kende.[46]
In de context van wetenschappelijk onderzoek lijkt de kans op zulke ongewenste spontane herkenning, althans bij grotere datafiles, uiterst onwaarschijnlijk. Het kan voorkomen dat iemand een individueel record met onderzoeksgegevens ziet en daarin een vroegere patiënt herkent. Dat zal dan zijn omdat deze onderzoeker eerder bij de behandeling betrokken was en de ziektegegevens die in dit record zijn opgenomen al kende. Dat de onderzoeker daarbij meer over deze patiënt zou kunnen vernemen dan hij of zij al wist of mocht vernemen, is niet goed voorstelbaar bij onderzoeksgegevens die overigens op grond van het aggregatieniveau als anoniem mogen worden beschouwd.

Dat lost naar mijn mening ook de vraag op of men bij de bepaling van het aggregatieniveau rekening moet houden met de kans op spontane herkenning door de onderzoeker. Bij een patiënt met een uiterst zeldzame aandoening zal spontane herkenning door de paar specialisten in het vak nooit kunnen worden uitgesloten, tenzij men de record zo anonimiseert dat deze voor onderzoek onbruikbaar wordt, bijvoorbeeld door de leeftijd geheel weg te laten. Bij zo'n record, die voor anderen dus wel onherkenbaar is – bijvoorbeeld door een ruime regio te benoemen als woonplaats (voor zover al een relevant onderzoeksgegeven) en een vijfjaarsklasse –, is het niet goed denkbaar dat degene die eerder die patiënt kende als behandelaar of bijvoorbeeld in het kader van een second opinion, meer verneemt dan hij al wist of op grond van diens professionele betrokkenheid mocht weten.

Een andere kwestie is dat soms de rollen van behandelaar en onderzoeker zo met elkaar zijn verwe-

ven dat, ook al worden de records geanonimiseerd, elke patiënt daarin door de onderzoeker zal worden herkend. Dan kan men uiteraard niet meer van onderzoek met anonieme gegevens spreken. Anonimisering kan dan nog steeds van belang zijn, bijvoorbeeld in verband met anderen die bij het onderzoek worden betrokken, maar dan als een privacybeschermende maatregel.

4.7 Gebruik van het BSN

Op grond van de Wet BSN in de zorg dient het burgerservicenummer (BSN) van de patiënt als uniek idenitificatiemiddel te worden gebruikt in het berichtenverkeer tussen zorgaanbieders en tussen hen en zorgverzekeraars. Artikel 24 Wbp staat er dan in de weg dat dit nummer voor een ander doel wordt gebruikt, tenzij dat bij AMvB zou zijn toegelaten.
Zo'n AMvB is er (nog) niet. Het BSN mag dus niet rechtstreeks in zorgonderzoek worden gebruikt, zelfs niet met toestemming van de betrokkene. Bij onderzoek met vrijwilligers (panel) mag men dus niet naar hun BSN vragen.

Voor het wel mogen gebruiken van het BSN bij wetenschappelijk onderzoek (ook bij het sociaal-wetenschappelijk onderzoek) is gepleit in een notitie van de KNAW en de FEDERA.[47] In gezondheidsrechtelijke kring blijkt over zulk gebruik grote aarzeling te bestaan.[48] De redenen zijn mij niet geheel duidelijk. De angst lijkt te zijn dat aldus mogelijk te grote databestanden te gemakkelijk kunnen worden gekoppeld. Dat lijkt mij echter geen probleem mits dat voor zinvol onderzoek gebeurt en met de nodige privacywaarborgen.
De facto gebeurt koppelen met een *versleuteld (gecodeerd)* BSN ten behoeve van wetenschappelijk onderzoek al, bijvoorbeeld bij het Parelsnoer Initiatief. In de in noot 15 genoemde brief van de Minister van VWS wordt gesteld dat de huidige wetgeving koppeling via een BSN met als exclusief doel wetenschappelijk onderzoek niet toestaat. Indien echter het BSN wordt versleuteld aan de bron, onder de voorwaarden voor gecodeerd anoniem, zoals eerder beschreven, is het geen BSN meer maar een uniek pseudoniem. De wetgeving verzet zich niet tegen zulke versleuteling en aldus speelt *indirect* het BSN toch een rol bij het creëren van een uniek codenummer.

Momenteel is dan ook de rare situatie ontstaan dat indirect via wel het BSN kan worden gekoppeld voor wetenschappelijk onderzoek, maar men vrijwilligers niet naar hun BSN mag vragen. Potentiële privacy problemen worden zo eerder vergroot dan verkleind. Van die deelnemers aan een panel zullen de antwoorden met hun toestemming met andere databestanden worden vergeleken waarin wel het BSN voorkomt. Als het panel het BSN zou mogen gebruiken kan dat zonder de kans op meerlingen en andere verwarring. Nu blijken zekere onzuiverheden steeds voor te komen. Zij moeten, zoals al opgemerkt, dan handmatig worden uitgezuiverd. In plaats van direct naar de patronen moet men eerst meer naar de personen terug dan met gebruik van het BSN voor het onderzoek nodig zou zijn.

5. DE TTP BIJ HET GEZONDHEIDSONDERZOEK

5.1 De TTP in het algemeen

In 1999 werd voor het gebruik van een TTP in de financiële wereld gesteld dat '*de*' TTP niet bestaat. Daarvoor waren de voorwaarden waaronder van een TTP mag worden gesproken nog onvoldoende uitgekristalliseerd en was haar verschijningsvorm niet eenduidig (Knoops 1999). Gelet op de geheel verschillende doelstellingen van een TTP in de financiële wereld en voor het gezondheidsonderzoek (zie hoofdstuk 1) lijkt die conclusie nog steeds op te gaan.
In de literatuur (Duthler 1998) is een aantal algemene voorwaarden geformuleerd voor de inzet van een TTP bij financiële transacties. Die zijn overeenkomstig toe te passen voor het gebruik van een TTP bij zorgonderzoek. Zij zijn:

- Betrouwbaarheid: de mate waarin de partijen die gebruik maken van de TTP zich kunnen verlaten op de dienstverlening en het informatiesysteem van een TTP.
- Onpartijdigheid: de TTP dient zich te onthouden van werkzaamheden die naar hun aard partijdigheid vereisen.
- Onafhankelijkheid: de TTP mag niet in functionele of financiële zin afhankelijk zijn van haar afnemers.[49]
- Deskundigheid: de TTP dient de zorgvuldigheid en bekwaamheid van een redelijk handelend vakgenoot te betrachten. De kennis aanwezig binnen de TTP dient te voldoen aan de 'state of the art', die kennis dient op peil te blijven en er dient voldoende ervaring te zijn binnen de TTP.
- Continuïteit: de TTP dient voldoende middelen te hebben om ook op langere termijn succesvol te kunnen opereren.
- Beveiliging: de TTP dient een formeel en schriftelijk gedocumenteerd veiligheidsbeleid op te stellen en te hanteren.

Voor een TTP bij gezondheidsonderzoek is men er dan nog niet. Er dienen ook voorwaarden te worden gesteld aan hoe de TTP de identiteit van betrokkenen maskeert. Zie daartoe sectie 4.3 over hetgeen het CBP heeft opgemerkt over versleuteling. In de daar genoemde situatie vindt deze plaats via een TTP.

5.2 Een voorbeeld van een TTP in de zorg

ZorgTTP[50] biedt diensten aan in het kader van de zorgfinanciering en verantwoording, bijvoorbeeld bij de risicoverevening tussen zorgverzekeraars. Op haar website geeft zij aan dat zij de wijze van pseudonimisering van een gegevensbestand steeds aan het CBP ter kennisgeving voorlegt.
ZorgTTP levert een methodiek waarmee persoonsgegevens aan de bron worden omgezet naar een pseudoniem. Daarbij is het proces zodanig ingericht dat het aangemaakte pseudoniem door de ontvanger niet langer kan worden herleid naar het persoonsgegeven waarop het pseudoniem is gebaseerd. Het is mogelijk om op basis van dezelfde persoonsgegevens bij andere bronnen hetzelfde pseudoniem (voor het betreffende domein) aan te maken. Hierdoor ontstaat de mogelijkheid om individuen in de tijd te volgen zonder dat daarvoor persoonsgegevens nodig zijn. Daarnaast is het mogelijk om pseudoniemen die afkomstig zijn uit verschillende domeinen (bronnen), maar hetzelfde individu betreffen, te koppelen.[51]
Overigens is ZorgTTP niet de enige organisatie die dergelijke diensten in Nederland aanbiedt. Het Parelsnoer-initiatief heeft na een Europese aanbesteding voor het bedrijf Diginotar gekozen.[52] Het zogenaamde Mondriaan-project[53] maakt gebruik van een Belgisch bedrijf.

5.3 De TTP bij gezondheidsonderzoek

Elk onderzoeksbestand kent twee typen gegevens:

a. de onderzoeksgegevens (dat wil zeggen de gegevens waarop de analyses worden verricht), en
b. het unieke pseudoniem voor elke set van onderzoeksgegevens die bepaalt dat deze onderzoeksgegevens bij één bepaald individu horen.[54]

Een TTP bij gezondheidsonderzoek biedt een vergelijkbare functie als de TTP bij de zorgverzekering en -administratie. Er is sprake van een keten van gegevens. Er is een verstrekker, zoals een behandelaar of een bepaalde registratie, die over de gegevens beschikt. Er is een ontvanger die met die gegevens onderzoek wil doen.
Als betrouwbare derde partij zorgt de TTP ervoor dat met betrekking tot b bij de verstrekker cryptografische technieken worden toegepast. Hierdoor komen de onderzoeksgegevens bij de ontvanger onder een 'b-versleuteld' aan, met andere woorden onder een pseudoniem. Via dit pseudoniem kan de ontvanger de verstrekte onderzoeksgegevens nog steeds aan bepaalde individuen koppelen. Voor de ontvanger zijn de set a en b echter anoniem, door zowel het aggregatieniveau van a als de versleuteling van b.

De TTP krijgt dus *geen* toegang tot het a-bestand (in combinatie met b). Anders is er niets gewonnen. Voor het b-bestand lijken er twee opties. In de meest sterke vorm draagt de TTP er zorg voor dat bij de verstrekker de cryptografische technieken worden toegepast. De TTP levert dan als het ware de software. Bij ZorgTTP wordt overigens na deze versleuteling aan de bron nog een tweede versleuteling toegepast. Versleuteling aan de bron is optie 1.
De tweede mogelijkheid is dat de TTP wel over de NAW-gegevens van de verstrekker beschikt, of daartoe toegang heeft en zelf versleutelt. Dat is optie 2.

Ad 1
De TTP verwerkt geen gegevens, dus ook geen persoonsgegevens. Uiteraard kunnen wel vragen worden gesteld over de beveiliging van het versleutelingsmechanisme bij de TTP. Zie daarover de volgende sectie.

Ad 2
Deze optie is in de praktijk naar voren gekomen, al wordt deze in de gepubliceerde stukken over een TTP bij de zorg niet genoemd. Zij wordt soms als een noodzakelijke tussenstap ervaren voor bestanden waarbij geen BSN is gebruikt. Bestanden vanuit meerdere verstrekkers dienen bij de onderzoeker te worden samengevoegd onder de unieke b. Deze verstrekkers kunnen echter verschillende annotaties hebben gebruikt voor de NAW-gegevens en er zullen altijd foutjes zijn gemaakt. De TTP beschikt over de techniek om deze eventuele dubbelingen uit te zuiveren. Er blijft dan een geschoond bestand over, waarop de versleuteling wordt toegepast.
Juridisch is er dan echter een probleem, want de TTP verwerkt wel persoonsgegevens. Bij zorgonderzoek is de afzender veelal een zorginstelling en zelfs als dan geen a-gegevens bij de TTP bekend zijn, maakt het loutere feit dat de NAW-gegevens komen van bijvoorbeeld een ziekenhuis, revalidatiecentrum of verpleeghuis deze al tot bijzondere persoonsgegevens.
Volgens de Wbp zou de TTP bewerker kunnen zijn voor de verstrekker. Die verstrekker blijft dan de verantwoordelijke. Vanuit die optiek zou deze techniek daarmee kunnen worden gelegitimeerd, mits verwerken van de gegevens ten behoeve van wetenschappelijk onderzoek ook in de doelomschrijving van de gegevensverwerking van de verantwoordelijke is opgenomen. De bewerker behoort, zoals in een andere Opinie van de artikel 29 Werkgroep is omschreven, tot de 'inner circle of data processing'.[55]
Vanuit de WGBO stuit deze optie op bezwaren. Patiëntgegevens worden verstrekt aan een partij die niet rechtstreeks bij de uitvoering van de behandelingsovereenkomst is betrokken. Indien dat *uitsluitend* gebeurt[56] ten behoeve van wetenschappelijk onderzoek kan dit in beginsel niet zonder toestemming van de betrokken patiënten. De Gedragscode Gezondheidsonderzoek voorziet hier niet in een uitzondering.[57]
De optie zou uitsluitend kunnen werken indien de door de TTP bewerkte NAW-gegevens niet met een bijzonder persoonsgegeven kunnen worden geassocieerd. Volgens artikel 9.3 Wbp mogen niet-bijzondere persoonsgegevens verder voor wetenschappelijk onderzoek en statistiek worden verwerkt mits die verwerking daartoe beperkt blijft. Als het gaat om de NAW-gegevens van bij huisartsen ingeschreven patiënten of bij zorgverzekeraars ingeschreven verzekerden, zou de TTP niet de beschikking krijgen over patiëntgegevens, ook niet indirect. Iedereen heeft – in beginsel –

een huisarts en een zorgverzekeraar. De huisarts brengt, door via een TTP de NAW-gegevens tot een pseudoniem om te werken, geen bijzondere persoonsgegevens naar buiten.

Bij zorginstellingen kan de TTP een dergelijke functie echter niet vervullen.
Overigens lijkt dit een tijdelijk probleem, namelijk tot het moment dat alle patiëntgegevens onder het BSN zijn genoteerd. Het zal dus voornamelijk spelen bij oude bestanden.
Voor het vervolg wordt optie 2 niet meegenomen.

5.4 Maar wat is nu de meerwaarde?

De TTP zorgt ervoor dat gegevens de onderzoeker gecodeerd (gepseudonimiseerd) bereiken. Maar dat zou de verstrekker ook zelf kunnen. De in sectie 3.4 beschreven casus van de CRF's naar de sponsor van een geneesmiddelenonderzoek is daarvan een voorbeeld. Een TTP komt daar niet aan te pas.
Maar in dat voorbeeld zijn er in elk centrum steeds verschillende patiënten. Er is, met andere woorden, geen patiënt die in twee centra aan dezelfde trial deelneemt. Bij de sponsor hoeven de onderscheiden deelnemers niet uniek te worden gekoppeld omdat elke deelnemer per definitie uniek is. Ieder centrum kan dus de eigen unieke codering toepassen.
De TTP lijkt dan met name aan de orde indien vanuit meerdere verstrekkers onderzoeksgegevens naar de onderzoeker worden gezonden die op dezelfde persoon betrekking hebben. Dan moet de codering worden gedeeld. De TTP beschikt over de techniek om deze codering bij de verschillende afzenders uniform te doen toepassen en aldus de stroom van onderzoeksgegevens onder het unieke pseudoniem (of codenummer) aan de verkrijger zodanig te organiseren dat deze bij dezelfde individuen kunnen worden samengevoegd. En dat gebeurt dan op een zodanige wijze dat de identiteit van de patiënt bij derden, waaronder de onderzoeker, geheim blijft.
Met andere woorden: de TTP zorgt op een aantoonbaar betrouwbare manier voor de anonimiteit van gegevens bij grotere onderzoeken, waarbij data uit verschillende bronnen moeten worden samengevoegd.

In de praktijk zal de TTP, om 'replay back' te voorkomen, bij een langerdurend onderzoek steeds nieuwe pseudoniemen aanmaken. De oude file met gecombineerde gegevens uit jaar 1 wordt dan opgevolgd door een nieuwe file met gecombineerde gegevens uit jaar 2 met nieuwe pseudoniemen.

5.5 De TTP als ultieme PET

PET staat voor 'privacy enhancing technologies'. Er bestaat een scala aan definities van PET, onder andere beschreven in Borking 2010. De Europese Commissie benoemt PET als 'a coherent system of ICT measures that protects privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data, all without losing the functionality of the information system' (Commission 2007). Niet alleen zoals gezien voor zorgonderzoek, maar bij alle gegevensverwerking wordt momenteel sterk ingezet op PET als middel om privacybescherming met processen van gegevensverwerking te verzoenen, zoals onder meer blijkt uit Commission 2010. Borking (2010, p. 196) noemt een aantal PET 'trappen'. Anonimiseren, zoals bij een TTP gebeurt, lijkt de hoogste te zijn. In die zin vormt inschakeling van een TTP de hoogste vorm van PET bij zorgonderzoek. Ook de PET definitie van de Europese Commissie benoemt als eerste 'verwijderen' van persoonsgegevens. De definitie eindigt echter met 'zonder de functionaliteit van het informatiesysteem in gevaar te brengen'. Of dat laatste voor onderzoekers ook opgaat, wordt in een volgend hoofdstuk onderzocht. Daarnaast kan de vraag worden gesteld of de koppelingen die via TTP mogelijk zijn en resulteren in grote databases zonder nadere voorwaarden nu wel recht doen aan de filosofische achtergrond van privacybescherming. Dat komt in het Engelse deel aan de orde.

5.5 Een ontnuchterende kanttekening

Die anonimiteit wordt bereikt door de combinatie van eigenschappen van de TTP, als het ware bekroond door de audit (zie secties 3.2 en 5.1). Toch zal ook dit systeem ervan moeten uitgaan dat bij TTP geen interne regels worden overtreden. Het versleutelingsmechanisme is daar bekend en, voor zover die er is, ook de sleutel terug.
Dit zou aan onderzoekers bekend kunnen worden gemaakt. De betrokken medewerker van de TTP zal uiteraard worden ontslagen en het vertrouwen in de TTP als onderneming zal een zeer forse deuk hebben opgelopen. Maar dat geldt ook voor de onderzoeker die indirect herleidbare personen in een onderzoeksbestand identificeert. Als dat uitkomt, volgt ontslag en, zo niet het einde van diens carrière, dan toch wel een forse terugslag. Het onderzoeksinstituut waaraan de onderzoeker was verbonden, zal aanzienlijk minder gemakkelijk gegevens van behandelaars of van patiënten zelf kunnen verkrijgen en veel moeten doen om het vertrouwen te herstellen.
Het moge nogal onorthodox zijn, maar het verschil tussen TTP en onderzoeker is minder groot dan wel wordt verondersteld.
Uiteindelijk komt het *ook* neer op procedurele regels binnen de TTP en het vertrouwen dat men zich daaraan conformeert. Dit kan bijvoorbeeld via logging achteraf worden gecontroleerd, maar nooit vooraf met 100% zekerheid worden afgedicht.

6. ALTERNATIEF VOOR DE TTP: PROBABILISTISCH KOPPELEN

Naast unieke koppeling via de NAW-gegevens of een op deze gebaseerd codenummer wordt in de onderzoekswereld op nog een andere manier zo uniek mogelijk gekoppeld, namelijk via het zogenaamde probabilistisch koppelen.[58] Deze techniek komt er in essentie op neer dat onderzoeksgegevens en een aantal niet indirect identificerende gegevens van de betrokkenen uit verschillende bronnen bij elkaar worden gevoegd. Die niet indirect maar wel enigszins identificerende gegevens zijn dan bijvoorbeeld geboortedatum, enkele letters van de naam, twee cijfers van de postcode. Via een aantal stappen wordt dan bepaald wie in deze pool van gegevens waarschijnlijk unieke individuen zijn. Deze krijgen een uniek nummer en de onderzoeksgegevens worden aan deze nummers gekoppeld. Er blijft altijd een groep gegevens over waarvan niet met zekerheid kan worden gesteld bij welke individuen deze horen, of, anders gezegd, of deze gegevens hoogstwaarschijnlijk op dezelfde unieke individuen betrekking hebben of op verschillende.

In termen van kosten en betrouwbaarheid voor het onderzoek is tussen de TTP en probabilistisch koppelen nog geen afweging gemaakt. Daartoe zou ook meer onderzoek nodig zijn dan voor deze notitie mogelijk was.
Probabilistisch koppelen is eveneens geen goedkope techniek, maar aanzienlijk goedkoper dan een TTP. Het kan voorts door de onderzoekers zelf worden uitgevoerd, met behulp van de statistische technieken die men daar van oudsher al in huis heeft. Probalistisch koppelen levert echter statistische waarschijnlijkheden op bij grote aantallen betrokkenen en is alleen voor bepaalde vraagstellingen bruikbaar. Voor het merendeel van het onderzoek geldt dat de onderzoekers moeten kunnen garanderen dat de gegevens tot die ene persoon behoren, wil het onderzoek betrouwbaar zijn.[59]

7 NADELEN VAN EEN TTP

7.1 De kosten

De TTP is een vrij gecompliceerde constructie. Daaraan zijn uiteraard kosten verbonden. De door een ICT verzorgde diensten moeten vanuit het onderzoeksbudget worden betaald. Het gaat dan met name om het inrichten van de gegevensstromen en de installatie van pseudonimiseringssoftware bij de verstrekkers. Daarbij moet al snel aan ten minste enkele tienduizenden euro's worden gedacht, los van de kosten van het jaarlijkse gebruik van de diensten.[60]

7.2 Het probleem van anonieme gegevens

7.2.1 Onderzoekers hebben vaak gedetailleerde gegevens nodig

Het uitgangspunt van de TTP is dat door deze de onderzoeksgegevens zodanig worden geaggregeerd[61] dat deze inderdaad anoniem zijn. Dit aggregatieniveau zal al snel te hoog zijn om met name bij zeldzame ziektes meer subtiele verbanden te kunnen onderzoeken tussen blootstelling, behandeling en de uitkomsten daarvan. Veel onderzoekers stellen dat ook buiten de context van zeldzame ziektes men een gedetailleerd inzicht nodig heeft in de behandeling van de patiënt (de episodes) om het behandelingsresultaat van de behandelingsinteracties goed te kunnen onderzoeken.
Natuurlijk hangt het van het type onderzoek af. Voor veel onderzoek kan men volstaan met meer globale gegevens. Met name dat onderzoek dat niet zozeer uitkomsten bij de patiënt onderzoekt, maar handelingen van hulpverleners. Maar om die patiënt is het uiteindelijk begonnen. Dan zijn op dat niveau gedetailleerde gegevens nodig om te kunnen bepalen hoe diens ziekteproces is verlopen en wat mogelijk aan de oorzaak of het niet genezen ten grondslag ligt dan wel juist aan de genezing heeft bijgedragen.

7.2.1 De uitdaging aan anonieme gegevens vanwege het aggregatieniveau

Het is de vraag of de gedetailleerde gegevens die voor het laatste type onderzoek nodig zijn, vanwege het aggregatieniveau nog wel als anoniem kunnen worden aangemerkt. Uiteraard kan die vraag uitsluitend worden gesteld vanuit de huidige opvatting, althans tenminste bij toezichthouders, dat onderzoekers het type recherchewerk zullen verrichten dat dan nog steeds nodig is om de identiteit van de patiënt te achterhalen.

Er is zelfs de vraag gesteld of überhaupt nog van anonieme gegevens kan worden gesproken. Ohm (Ohm 2010) heeft laten zien dat het al snel mogelijk is om de betrokkenen achter anonieme gegevens te identificeren door die ogenschijnlijk anonieme gegevens via overigens vrij geavanceerde statistische analyses te koppelen met openbare gegevens van personen op het internet. In vergelijkbare zin werd daarom door Korff kritiek geuit op meergenoemde Opinie 4/2007. De Werkgroep zou te veel de nadruk hebben gelegd op de pseudonimisering van de direct identificerende gegevens en te weinig aandacht hebben gehad voor de herleidbaarheid van de gegevens die vervolgens overblijven (Korff 2010).
De door Ohm genoemde voorbeelden betreffen overigens niet gegevens uit het onderzoekdomein, waarvan in de regel minder corresponderende gegevens ook op persoonsniveau op het internet te vinden zijn. Maar de waarschuwing blijft, anonieme gegevens worden een steeds beperktere categorie. Ohm concludeert in dat kader zelfs tot het 'failliet' van de Europese sectorbrede en technieknetrale benadering zoals volgt uit Richtlijn 95/46/EC.
Of dit laatste nu helemaal waar is of niet, het feit blijft dat bij de huidige opvatting over gegevens binnen het onderzoekdomein, steeds minder gegevens als anoniem kunnen worden aangemerkt. Een TTP zal dus steeds minder een oplossing bieden voor het probleem van toestemming of geen bezwaar om gegevens voor zorgonderzoek te mogen gebruiken.
Op de consequenties van deze conclusie kom ik terug in het Engelstalige deel.

7.3 Eenweg codering

Het derde nadeel is dat een TTP in beginsel uitsluitend 'eenweg' codeert. Vaak willen onderzoekers echter over bepaalde patiënten meer weten. Bijvoorbeeld wanneer diagnose en medicatie niet overeenkomen of wanneer men zich realiseert dat men voor een goed begrip van de behandeling of de effecten bij deze patiënt extra gegevens nodig heeft. Omdat de codering eenweg is, kan de onderzoeker echter niet via de TTP bij de bron nadere inlichtingen over bepaalde patiënten nadere inlichtingen vragen.

In het algemeen wordt niet beseft hoe 'ambachtelijk' veel van dit onderzoekswerk is. Ik dacht daarover bij de start van mijn onderzoek ook veel te naïef. Ook bij grotere databestanden kent het onderzoek meestal 2 stappen nadat de gegevens in het onderzoekdomein zijn gearriveerd. Eerst wordt record voor record gecontroleerd om te bepalen of de gegevens wel voldoende accuraat zijn om als onderzoeksgegevens te worden meegenomen. Pas dan gaan ze in de database van de onderzoeksgegevens of worden ze daarin aangemerkt als voldoende valide om mee te tellen. De eenwegcodering bij de TTP maakt het echter onmogelijk om de eerste stap goed uit te voeren, dus over bepaalde onzuivere records nadere vragen bij de bron te stellen.

7.4 Het oude bestand wordt vervangen door een nieuw bestand (met nieuwe pseudoniemen)

Bij longitudinaal onderzoek via een TTP wordt het bestand uit 'jaar 1' in 'jaar 2' vervangen door een nieuw bestand met nieuwe pseudoniemen. Over het bestand uit jaar 1 zal men in het algemeen niet meer mogen beschikken, om het gevaar van 'replay back' te voorkomen. Voor al die verzamelingen die tot niet meer leiden dan statische overzichten, zoals wachtlijstbeheer, kan zo'n systeem goed werken. Maar onderzoekers zullen wel over die oude bestanden willen kunnen beschikken. Al is het maar omdat analyses en publicaties meestal niet binnen één jaar zijn afgerond. Maar ook als dat wel het geval zou zijn, die resultaten moeten altijd kunnen worden gereproduceerd. Onderzoekers moeten dus gedurende langere tijd over de oorspronkelijke onderzoeksgegevens kunnen beschikken.
Dat dit 'replay back' via de TTP moet worden voorkomen heeft dezelfde achterliggende reden als eerder genoemd, namelijk de vooronderstelling van fraude door onderzoekers. In dit geval niet door recherchewerk om indirect identificerende gegevens te achterhalen, maar door tegen de afspraken in via vergelijking van bestanden de pseudonimiseringsprocedure te ontrafelen.

7.5 Onzuivere 'regulation by technology'

Daarmee kom ik op het vijfde nadeel, dat van een geheel andere orde is. Een gecompliceerde constructie wordt ingezet voor wat wel beschouwd een probleem van vertrouwen is. De gedachte dat onderzoekers eerder genoemd recherchewerk zouden toepassen om indirect identificerende gegevens te achterhalen of 'replay back' proberen, is – netjes gezegd – wel heel erg vergezocht. Ik noemde in sectie 4.2 al de reacties van onderzoekers op de kans op het doorbreken van de interne coderingen.
In een ander verband heeft Bronsword de inzet van techniek om gedrag te normeren 'regulation by technology' genoemd (Bronsword 2008). Hij gaat een stap verder. De normatieve afwegingen die actoren zouden moeten maken worden hen onthouden en vervangen door een technologisch keurslijf. Een volledige bespreking van deze stelling zou buiten het bestek van deze notitie vallen.
Hier is wel het volgende van belang. De voorbeelden van Bronsword, zoals een 'alcoholslot' op auto's, betreffen situaties waarin het bepaald onzeker is of iedereen wel de juiste afwegingen maakt en in ieder geval gebleken is dat sommigen die niet zullen maken.
De situatie van het zorgonderzoek verschilt hier hemelsbreed van, althans bij de onderzoekers in COREON-verband, met wie ik in en buiten de context van deze notitie veelvuldig heb overlegd. Van 'lekken' buiten het onderzoekdomein[62] is nog nooit gebleken (en dat zou zeker aan het licht zijn gekomen). Dat binnen het onderzoekdomein onderzoekers meer doen dan nodig is voor het herkennen van patronen (in plaats van personen),[63] valt uiteraard lastiger aan te tonen, maar zelfs

anekdotisch zijn daar mij in ieder geval geen voorbeelden van bekend.[64]

Technologische constructies als een TTP bestaan dankzij een klimaat van georganiseerd wantrouwen, naast andere factoren zoals het idee dat risico's in alle opzichten moeten worden afgewend. Die tendens kan onder andere leiden tot disproportionele inperking van persoonlijke vrijheid en van de normativiteit die van actoren zelf mag worden verwacht.
Mogelijk is dat in bepaalde maatschappelijke 'contexts'[65] onvermijdelijk. De risico's zijn hoog, de kans op niet norm conform gedrag om die risico's af te wenden eveneens. Wellicht zou ook de maatschappelijke winst van de desbetreffende activiteit een rol kunnen spelen in de balans.
In ieder geval moet het gaan om een genuanceerde benadering. In dit geval één die recht doet aan de belangen van zorgonderzoek voor de gezondheidszorg, dus voor patiënten, in samenhang met de daarbij noodzakelijkerwijs te gebruiken technieken en de vraag welke onderzoekers het onderzoek uitvoeren en de af te wenden risico's. In het Engelstalige deel ga ik hier nader op in.

8. DE TTP IN HET SPECTRUM VAN PET

8.1 Inleiding

Het doel van dit hoofdstuk is om enkele globale aanwijzingen te geven wanneer welke PET bij gezondheidsonderzoek gewenst is, uitgaande van de huidige regelgeving en de interpretatie van deze waarbij de vooronderstelling van kwade trouw min of meer het uitgangspunt is. Op een bepaald moment loopt die benadering vast. Het streven naar anonieme gegevens in het onderzoekdomein leidt tot een gegevenswoestijn waarin goed onderzoek onmogelijk is. In het Engelstalige deel kom ik daarop terug en geef ik aanwijzingen hoe dit vastlopen kan worden voorkomen.

Binnen de huidige benadering kan het volgende worden opgemerkt. De Gedragscode Zorgonderzoek van 2004 gaat er al van uit dat persoonsgegevens zo min mogelijk worden gebruikt. Het in de Gedragscode aanbevolen onderscheid tussen een bestand met onderzoeksgegevens en een communicatiebestand is al een vorm van PET. Uit het voorgaande is duidelijk dat er echter veel meer moet gebeuren om dat bestand met onderzoeksgegevens als anonieme gegevens te mogen aanmerken. Zoals opgemerkt, het inschakelen van een TTP kan als de meest vergaande vorm van PET worden beschouwd. Er zijn talloze tussenvormen en deze meest vergaande vorm lijkt niet altijd nodig te zijn. Temeer daar er ook de net behandelde nadelen aan zijn verbonden. Zij *kan* zelfs niet altijd worden gebruikt zonder aan de functionaliteit van het systeem afbreuk te doen, indien dat wordt gezien als een systeem beginnende met de bronnen van patiëntgegevens tot en met gegevens in het onderzoekdomein die voor zinvol, genuanceerd zorgonderzoek voldoende valide zijn.

8.2 PET in het licht van de zwaarte van de risico's

Het bovenstaande betekent evenmin dat uitsluitend vertrouwen in onderzoekers de basis zou moeten zijn voor de verstrekking van de primaire bronnen naar onderzoekers en de gegevensverwerking binnen het onderzoeksdomein. Op dat 'vertrouwen' en hoe dat binnen het onderzoekdomein te borgen ga ik nader in in het Engelstalige deel.
Hier gaat het om een aanzet voor de beoordeling van de inzet van een TTP in de zin van de proportionaliteit van de inzet van deze PET. De voorwaarden die aan PET's moeten worden gesteld, dienen in verhouding te staan tot de mogelijkheden van patiënten om de gegevensverwerking te beïnvloeden, zoals bepaald door het geldende recht, de risico's van de gegevensverwerking maar tevens in het licht van de mogelijkheden om nog zinvol zorgonderzoek te kunnen verrichten.

PET in de zin van het slechts dan gebruiken van persoonsgegevens als dat voor het doel van het onderzoek strikt nodig is en geoorloofd is op basis van de toepasselijke regelgeving, moeten altijd worden toegepast. Er is geen onderzoeker - althans voorzover ik dezen heb gesproken – die dit uitgangspunt ontkent.
De vraag is hoe ver de daartoe ingeroepen technieken dan moeten gaan. Dat hangt af van een aantal criteria, die elkaar deels overlappen:

a. De belangrijkste is of er al dan niet toestemming is verleend voor gebruik van persoonsgegevens voor wetenschappelijk onderzoek. Indien dat niet het geval is, zal een sterke vorm van PET moeten worden gebruikt. Onderzoekers zullen immers slechts anonieme persoonsgegevens mogen gebruiken, tenzij één van de uitzonderingen van de Gedragscode Gezondheidsonderzoek aan de orde is. Zoals gezien (hoofdstuk 2) bieden deze uitzonderingen in veel gevallen onvoldoende soelaas.
   Anders dan de andere, hierna volgende criteria die gradueel meer of minder wegen, is criterium a als het ware 'quartair'. Het is van de volgende vier mogelijkheden:
   - Er is toestemming voor het gebruik van persoonsgegevens voor onderzoek.
   - Er is geen toestemming voor, maar evenmin geen bezwaar tegen het gebruik van zulke persoonsgegevens in uitzonderingssituaties.
   - Er is geen toestemming voor, maar wel bezwaar tegen gebruik van persoonsgegevens voor wetenschappelijk onderzoek.

- Er is geen toestemming, het is onbekend of er geen bezwaar is.

Deze laatste twee modaliteiten vallen voor de regelgeving samen. Uitsluitend anonieme gegevens mogen dan worden gebruikt.

b. De complexiteit van de gegevensverwerking en (veelal in samenhang daarmee) het aantal daarbij betrokken onderzoekers en medewerkers: naarmate die complexiteit groter is en de kans groter is dat anderen dan de direct betrokken onderzoekers van persoonsgegevens kennis kunnen nemen, zullen PET's die kans zoveel mogelijk moeten verminderen. Dat geldt daarmee ook indien ten behoeve van het onderzoek in beginsel wel persoonsgegevens mogen worden gebruikt.
c. Tijdsduur van het onderzoek: naarmate deze langer is, is er reden om de bestaande procedures en werkwijzen op een bepaald moment tegen het licht te houden, zowel op de conformiteit van de werkwijzen met de procedures als op de effectiviteit van de procedures zelf. Bijvoorbeeld indien herhaaldelijk ten behoeve van hetzelfde onderzoek van dezelfde patiënten gegevens versleuteld worden aangeleverd, zal de procedure op gezette tijden moeten worden herzien, eventueel een nieuwe wijze van versleuteling moeten worden ingevoerd.
d. De mate van gevoeligheid van de gegevens. Alle gegevens vanuit de zorg zijn gevoelig maar sommige zijn meer gevoelig dan andere. Het maakt nogal een verschil of men nagaat wie in het weekend een spoedeisende hulp heeft bezocht in verband met een sportletsel of voor wie in het weekend tot een gedwongen opname in een psychiatrisch ziekenhuis is besloten.
e. De omvang of hoeveelheid gegevens is tot slot eveneens een factor waarmee rekening moet worden gehouden. De gegevens van slechts enkele patiënten die bij één onderzoeker worden gehouden, verdienen een andere wijze van beveiliging dan een via het internet toegankelijke database met alle patiënten, al dient het resultaat in termen van gegevensbescherming uiteraard gelijk te zijn.[66]

8.3 Afweging PET in het licht van doelmatigheid?

Impliciet in de bovenstaande benadering is dat de zwaarste PET's niet altijd hoeven te worden ingezet. De facto gebeurt dat ook niet, het onderzoek zou dan volstrekt onbetaalbaar zijn geworden. Kleinere onderzoeken zouden onmogelijk zijn. Maar het is goed om dit uitgangspunt expliciet te benoemen, nu de zeer grote onderzoeken, opgezet met aanzienlijke voor de kennisinfrastructuur bedoelde financiële middelen, de norm lijken te worden. Zulke onderzoeken scoren hoog op alle genoemde criteria.
Er vindt echter zeer veel onderzoek plaats waarbij dit niet het geval is, bijvoorbeeld panelonderzoek waarbij er per definitie wel toestemming is. Het onderzoek scoort dan 0 op criterium a. Onder omstandigheden kan deze toestemming echter gebrekkig zijn in het licht van de huidige mogelijkheden terwijl je niet meer naar de deelnemers terugkunt, bijvoorbeeld omdat dezen zijn overleden. Dan moet eventueel op de tweede modaliteit (geen expliciete toestemming, evenmin bezwaar) op dat criterium worden teruggevallen. Dit vereist evenwel niet steeds dat de gehele gegevensverwerking onder alle omstandigheden anoniem moet zijn.
Wel zal ook dan zoveel mogelijk een scheiding moeten worden aangebracht tussen contactgegevens en onderzoeksgegevens. Uiteraard zal ook de externe veiligheid moet zijn geborgd. Anderen dan de daartoe aangewezen onderzoekers mogen geen toegang hebben tot de gegevens.

En zo zijn er veel meer nuances, afhankelijk van hoe het onderzoek is opgezet en daarmee op de genoemde criteria scoort.
Als men wil dat publieke middelen doelmatig worden gebruikt, geldt dat ook voor die voor zorgonderzoek. Deze zijn toch al relatief schaars. TTP- en auditconstructies zijn, hoe fraai ook, kostbaar. Zij zouden uitsluitend dan moeten worden ingezet wanneer of de techniek van het onderzoek of de bijzondere aan dit onderzoek verbonden risico's dat vereisen. Zij zouden voorts additioneel moeten zijn op wat onderzoekers zelf al hebben geregeld. Daarop zal in het Engelstalige deel worden teruggekomen.

8.4 Altijd een audit bij criterium a, geen toestemming?

In het verlengde van de opmerking uit de vorige sectie het volgende.
Zoals gezien eiste het CBP een audit bij de gegevensverstrekking aan het CVZ. De vraag is of daaruit moet worden afgeleid dat een dergelijke audit altijd noodzakelijk is bij het verstrekken van onderzoeksgegevens onder een codenummer.
Deze door het CBP beoordeelde gegevensverwerking scoort ook hoog op alle andere genoemde criteria. Daarbij komt nog eens dat het CVZ een orgaan is dat dicht bij de overheid is gepositioneerd, waar ook gegevens in verband met de risicoverevening in het kader van de Zvw worden verwerkt en dat adviseert over zorgaanspraken.
Feit is dat thans zo'n audit bij veel gecodeerde gegevensverstrekking van bronbestanden aan onderzoekers niet plaatsvindt. De kosten van onderzoek zouden exponentieel stijgen indien dat wel zou moeten gebeuren. Behalve auditors en eventueel toezichthouders zou niemand daar wel bij varen. Het lijkt mij dat ook de criteria b en c, een rol moeten spelen of dergelijke audit opportuun is. Dit naast een additioneel criterium dat hiervoor al werd geïmpliceerd:

f. Welke veiligheidsmaatregelen hebben onderzoekers zelf genomen om de rechtmatigheid en veiligheid van gegevensverwerking bij het onderzoek te borgen?

Zoals opgemerkt wordt hierop in het Engelstalige deel teruggekomen.

8.5 Wanneer nu een TTP?

Voor een onderzoek dat hoog scoort op alle genoemde criteria zal een TTP een oplossing kunnen bieden voor de volgende samenhangende problemen:

1. Borging dat gegevens van over bepaalde personen uit het bronsysteem van de verstrekker de onderzoeker onder een uniek pseudoniem bereiken.
2. Borging dat gegevens vanuit verschillende verstrekkers die betrekking hebben op dezelfde individuen, de onderzoeker onder hetzelfde pseudoniem bereiken;
3. Verantwoording van de betrouwbaarheid van dit proces tegen derden.

Echter, uitsluitend dan indien

4. De in de paragraaf genoemde nadelen van een TTP voor het onderzoek geen onoverkomelijke problemen vormen.

Aan voorwaarde 4 zal lang niet altijd kunnen worden voldaan. Het zou zelfs onmogelijk zijn om er aan te voldoen indien Ohm's analyse van vrijwel altijd mogelijke herleidbaarheid van genuanceerde gegevens serieus moet worden genomen.

8.6 TTP als niet ultieme PET

Voor het onderzoekdomein lijkt mij de conclusie gerechtvaardigd dat de illusie van de TTP als ultieme PET, leidend tot anonieme gegevens in het onderzoekdomein, moet worden verlaten. Dat betekent zeker niet het einde van de TTP. Een TTP kan nog steeds worden ingezet. Vaak moet dat zelfs. Maar de TTP levert dan indirect herleidbare gegevens aan. Eventueel kunnen of moeten die zelfs tweeweg gecodeerd zijn.
De TTP wordt daarmee een 'gewone' PET die in een aantal omstandigheden bijzonder nuttig kan zijn. Een TTP kan data uit verschillende bronnen uniek linken in het onderzoekdomein. Mogelijk kunnen onderzoekers die technieken ook ontwikkelen, in combinatie met het datacenter dat aan veel UMC's is verbonden. Het is de vraag of dat kosteneffectief is. TTP's hebben die technieken al in huis.
Voorts speelt in de praktijk vaak het volgende. TTP's hebben vaak ook al pseudonimisering toegepast voor databases waar onderzoekers gegevens willen ontlenen. Deze databases hebben de gegevens via de TTP gepseudonimiseerd ontleend aan primaire bronnen. De onderzoekers willen gegevens uit de database combineren met gegevens uit andere primaire bronnen. Het is dan logisch dat het pseudonimiseringsproces leidend naar gegevens in het onderzoekdomein over de hele linie hetzelfde is en dus de TTP die de eerste pseudonimisering heeft toegepast, wederom

wordt ingeschakeld. Dit leidt overigens wel tot de vraag in hoeverre zo'n TTP niet een monopolie positie verwerft. Daarop aansluitend: of niet procedure moeten worden ontworpen dat TTP's onder voorwaarden elkaars pseudonimiseringsproces overnemen. Dat valt buiten het bestek van dit onderzoek.

Overnemen van elkaars pseudoniemen is zonder meer een potentieel veiligheidsrisico. Het zou niet passen in de idee van de TTP als ultieme PET. Het past evenwel in de idee van de TTP als een van de mogelijk in te zetten PET waarmee een optimale balans wordt gevonden tussen privacybescherming en de noodzaak van patiëntgegevens voor zorgonderzoek.
Die balans moet steeds worden gezocht. In het Engelstalige slothoofdstuk doe ik daartoe voorstellen.

8.7 De TTP in het licht van de uitzonderingen om gegevens te gebruiken zonder toestemming

De conclusies voor deze paragraaf zijn duidelijk. In een aantal gevallen zal de TTP inderdaad een oplossing kunnen bieden om patiëntgegevens niet op basis van de uitzonderingen naar het onderzoekdomein te leiden. Zij zijn dan anoniem en toestemming en geen bezwaar zijn naar huidig recht niet van toepassing.
Dat aantal gevallen is echter beperkt. De uitzonderingen zijn naar huidig recht nog steeds nodig.

Notes to the Dutch part

[1] Met klinisch onderzoek wordt gedoeld op onderzoek waarbij patiënten in het kader van wetenschappelijk onderzoek een bepaalde behandeling wordt aangeboden. Dat is een vorm van interventie-onderzoek en altijd onderzoek in het kader van de Wet medisch-wetenschappelijk onderzoek met mensen (WMO). Hiertegenover staat observationeel onderzoek zoals in deze notitie aan de orde is. Onder observationeel onderzoek valt ook het actief bevragen van personen naar hun leefstijl, ervaringen etc. Zulk observationeel onderzoek is in deze Notitie niet aan de orde. Het gaat in deze Notitie uitsluitend om onderzoek met reeds bestaande gegevens.

[2] Zie Porter 1997, p. 412 e.v.

[3] Veel voorbeelden bijvoorbeeld in Lowrance 2002.

[4] Voor 'gewone' persoonsgegevens geldt artikel 9.3 Wbp. Patiëntgegevens zijn echter 'bijzondere persoonsgegevens' in de zin van de Wbp en dan geldt artikel 23 Wbp.

[5] Stcrt. 2004, nr. 82.

[6] Federatie van medische wetenschappelijke verenigingen, zie www.federa.org.

[7] Zie bijvoorbeeld Van Veen 2008b

[8] Zie ook Stolker 2003. In deze opvatting past naar mijn mening dan ook slecht een voetnoot met 'anders P. Modaal, etc.', zoals in veel juridische publicaties helaas wel gebeurt. Als dat anders de kern van het betoog betreft, moet men dat andere (proberen te) weerleggen.

[9] Duthler 1998.

[10] Wmg, Zvw, AWBZ.

[11] In de zin van de Wbp, dat wil zeggen degene op wie een persoonsgegeven betrekking heeft.

[12] Het voorwoord van J. Kohnstamm bij CBP 2005.

[13] Registratiekamer 1995/2000.

[14] Voor een verslag zie www.privacyandresearch.be (laatst bezocht januari 2011). Idem met betrekking tot 'public health monitoring Carinci 2011.

[15] Brief van 4 december 2009, Kamerstukken II 2009/10, 27 543, nr. 9, p. 3.

[16] Zoals opgemerkt is de verklaring van overeenstemming door het CBP ex art. 24 Wbp van deze Gedragscode inmiddels verlopen. De regelgeving waarop de Gedragscode (WGBO, Wbp) is gebaseerd, is echter ongewijzigd gebleven. De Gedragscode vormt daarmee nog steeds een rechtmatige uitwerking van deze regelgeving, althans voor die onderwerpen die de Gedragscode wel regelt. Anders dan in de regeringsnota naar aanleiding van de Wbp-evaluatie wordt gesteld (Kamerstukken II 2009/10, 31 051, nr. 6) is het ook een inhoudelijk rijke Gedragscode die meer doet dan een opsomming geven van de relevante termen uit de Wbp en WGBO. De Gedragscode wordt thans herzien en zal ook antwoorden moeten bieden op vragen waarvoor onderzoekers thans staan, zoals betreffende de primaire data verzameling bij respondenten of het al dan niet inschakelen van een TTP.

[17] Voor deze terminologie en de wijze van toetsing zie Van Veen 2008c, met name p. 37.

[18] Een METC-beoordeling is niet nodig voor het inzien van enkele dossiers, teneinde te kunnen nagaan of de daarin gevonden informatie zich leent om daaruit gegevens voor onderzoek te genereren en dus een onderzoeksprotocol op te stellen.

[19] Beschreven bij Ingelfinger 2004.

[20] Zie verder sectie 10.3.4

[21] Appels 2007, verder de literatuur aangehaald bij Van Veen 2008a, Hansson 2009, zie verder sectie 10.3.4.

[22] Het voldoen aan NEN norm 7510 is dwingend voorgeschreven via een ministerieel besluit op grond van de Wet gebruik burgerservicenummer in de zorg. De norm wordt momenteel herzien.

[23] Dat geldt ook voor het geen bezwaar-systeem bij 'nader gebruik' van lichaamsmateriaal.

[24] Meer hierover Lowrance 2002.

[25] Vergelijk RGO 2008.

[26] Met 'aggregatieniveau' wordt gedoeld op de mate van detaillering van de onderzoeksgegevens die aan het unieke individu zijn gekoppeld. Geslacht en leeftijd (in verschillende varianten, bijvoorbeeld ten tijde van optreden ziekte, ten tijde van overlijden, etc.) zijn altijd relevante onderzoeksgegevens. Veelal geldt dat ook voor woon-of werkplaats, beroep, en natuurlijk ook de aandoening, etc. In combinatie kunnen die onderzoeksgegevens indirect identificerend zijn. Een hoger aggregatieniveau betekent dat men in plaats van leeftijd een jaarklasse neemt, in plaats van vier cijfers van de postcode slechts drie of twee, etc. Op de betekenis van het

aggregatieniveau wordt later teruggekomen.

[27] In Dute 2008 (p. 224, 225). wordt in het kader van de bespreking van artikel 7:458 BW gesteld dat het hier om gecodeerde gegevens gaat en bij de bespreking van de Wbp dat gecodeerde gegevens geen persoonsgegevens zijn. Dat zou misverstanden kunnen wekken. Gecodeerde anonieme gegevens over patiënten zijn geen patiëntgegevens in de zin van de WGBO net zoals zij geen persoonsgegevens zijn in de zin van de Wbp. Niet alle gecodeerde gegevens zijn echter anonieme gegevens (zie verder par. 4.3).

[28] De bepaling is bij nota van wijziging in het wetsontwerp op de geneeskundige behandelingsovereenkomst ingevoegd. In de aanloop naar deze bepaling was er een ietwat verhitte discussie in de media ontstaan of (in de huidige nummering) de artikelen 7:457 en 7:458 BW, met uitsluitend het in onderdeel 1a bepaalde, het wetenschappelijk onderzoek met patiëntgegevens niet onmogelijk zouden maken. Een artikel in het NRC (Van Veen 1994) leidde tot een opening. Onderzoekers werden uitgenodigd in de Kamercommissie die de zorgen van de onderzoekers deelde. De indieners kwamen met een nota van wijziging waardoor ook onderdeel 1b aan art. 7:458 werd toegevoegd. In het NRC-artikel werd gesproken van 'gecodeerde gegevens' terwijl art. 7:458 dit begrip vermijdt. Maar mogelijk dat hierdoor in een deel van het gezondheidsrecht het misverstand is ontstaan dat alle gecodeerde patiëntgegevens dus gegevens zijn in de zin van art. 7:458 lid 1 onder a, zijnde gegevens waarbij 'herleiding redelijkerwijs wordt voorkomen' en die dus nog wel persoonsgegevens zijn.

[29] Zie Overweging 26 bij deze Richtlijn.

[30] Zie ook Lowrance 2002 en hierna, Working Party 2007b.

[31] Deze voorbeelden zijn ontleend aan CBP 2005.

[32] Dat bleek ook tijdens de bespreking van versie 5 van deze notitie op de vergadering van Coreon door onderzoekers die niet bij deze casus waren betrokken. De verontwaardiging was algemeen en zelfs vrij heftig. Een andere kwestie is dat men dit helaas niet hardop tegen de toezichthouder durft te zeggen. Die heeft de macht en met die moet men verder. Om die reden werd ik ook verzocht deze passage af te zwakken. Hoewel men inderdaad verontwaardigd is, is men bang dit te verklaren. Dat zou de deur naar verder 'onderhandelen' kunnen sluiten. Het voordeel van het onder eigen naam publiceren van deze notitie is dat ik zulke verstandige opmerkingen zelf kan wegen en een mogelijk onverstandige, maar wel feitelijk juiste en duidelijke, kan laten staan. Dit zijn feiten die voor het verhaal van het recht van belang zijn.

Deze schroom tegenover de toch vrij machtige toezichthouder zou overigens ook wel eens een geheel andere reden kunnen hebben dan genoemd in het evaluatierapport van de Wbp, namelijk dat over de Wbp zo weinig wordt geprocedeerd. Niet omdat de betrokkenen hun rechten onvoldoende kennen maar omdat verantwoordelijken (of semi-verantwoordelijken, want men is geen verantwoordelijke in de zin van de Wbp als het geen persoonsgegevens betreft) het niet durven. Buiten de context van de bespreking van het positieve recht is deze notitie ook een pleidooi voor meer assertiviteit van onderzoekers dus vandaar deze opmerking in een voetnoot.

[33] Ploem 2004, p. 230 e.v. Idem Ploem 2010.

[34] Van Veen 2003; Quathem 2005, p. 155-161; Dute 2008, p. 225, althans voor gegevens in de zin van de Wbp.

[35] Deze is ingesteld op grond van artikel 29 van Richtlijn 95/46/EC en bestaat uit afgevaardigden van de nationale toezichthouders, zoals 'ons' CBP.

[36] Deze was oorspronkelijk bedoeld om tot geschoonde wachtlijsten te komen voor zorgaanspraken op grond van de AWBZ; met name die voor de gehandicaptenzorg bleken sterk vervuild doordat aanvragers of hun vertegenwoordigers zich bij meerdere instellingen hadden ingeschreven. Er was dus behoefte aan een unieke koppeling op één plek om te zien hoeveel mensen nu werkelijk wachten op een plaatsing in een instelling voor gehandicaptenzorg. Het diende op het niveau van het CVZ echter wel om anonieme gegevens te gaan.

[37] Zie echter Van Veen 2003.

[38] Working Party 2007c, p. 19.

[39] Hierop worden chronologisch basisgegevens over de diagnose, de gegeven medicatie en de gezondheid van de proefpersoon aangetekend.

[40] Bij dubbelblind onderzoek is er daarnaast nog een interne codering, waaruit blijkt of een patiënt de trial medicatie heeft gekregen of bijvoorbeeld het placebo. Voor het hier behandelde principiële punt van verstrekking van gecodeerde CRF's aan de sponsor doet dat niet terzake.

[41] Los van de andere GCP-eisen zoals geformuleerd in Richtlijn 2001/20/EC en daarop gebaseerde uitvoeringsrichtlijnen.

[42] Vergelijk ook de Mededeling van de Centrale Commissie Mensgebonden Onderzoek over codering van onderzoeksgegevens van 19 oktober 2009, www.ccmo-online.nl/main.asp?pid=25&sid=49&ssid=174 (laatst

bezocht mei 2011).

[43] Voor de goede orde. Dit is een ander soort onderzoekers dan de onderzoekers bij gezondheidsonderzoek. De in de tekst bedoelde onderzoekers zijn tevens behandelaars die in het kader van een medisch-wetenschappelijk onderzoek in de zin van de WMO een nieuwe mogelijkheid voor de behandeling toepassen. Bij gezondheidsonderzoek gaat het om onderzoekers die de betrokken patiënten meestal nooit zien maar uitsluitend de gegevens die uit het dossier van de patiënt (en eventueel andere bronnen) zijn ontleend, gebruiken om daaruit conclusies te trekken. Deze onderzoekers doen louter observationeel onderzoek op basis van elders beschikbare gegevens of op basis van door de deelnemer aan een onderzoek zelf gerapporteerde gegevens over diens omstandigheden (die verder niet worden gewijzigd).

[44] Dat geldt ook voor de METC en de Inspectie voor de Gezondheidszorg. Via informed consent geeft de proefpersoon hiervoor eveneens toestemming.

[45] O.a. bij Ploem 2010.

[46] Deze casus is niet verzonnen. Het betrof een college psychologie.

[47] Brief van 19 februari 2008.

[48] Dute 2008, Ploem 2010.

[49] Uiteraard is er altijd een financiële relatie in de opdrachtgever-opdrachtnemer (TTP) relatie. Bedoeld zal zijn dat de TTP niet uitsluitend van deze opdrachtgever afhankelijk mag zijn.

[50] www.zorgttp.nl.

[51] Voor een uitvoerige beschrijving van de inzet van een TTP bij de zorgverzekering zie Stichting Informatievoorziening Zorg, Het gebruik van pseudo-identiteiten binnen de risicovereveningssystematiek, Houten: 20 juni 2006.

[52] Nieuwsbrief Parelsnoer Initiatief, december 2009.

[53] De Stichting Mondriaan is een door het Topinstituut Pharma gesubsidieerd initiatief dat kort gezegd als 'datamakelaar' tussen organisaties in de zorg met gegevens (beroepsbeoefenaren, zorginstellingen, zorgverzekeraars) en onderzoekers wil bemiddelen.

[54] Naar aanleiding van een opmerking van een onderzoeker bij een eerdere versie het volgende. Een TTP is uitsluitend dan aan de orde indien de onderzoeker geen contact heeft met de betrokkene. In het kader van de TTP is er een keten van verstrekker naar onderzoeker. De onderzoeker kan de betrokkenen niet zelf benaderen. Bij het zogenaamde panelonderzoek door de onderzoeker is er wel contact, ook al vindt dat niet rechtstreeks plaats. Als PET wordt een tussenliggend bestand gemaakt, met NAW-gegevens waarmee de respondenten worden benaderd. Dat bestand is intern bij het onderzoeksinstituut gescheiden van het onderzoeksbestand. Administratieve medewerkers werken met het communicatiebestand, de onderzoekers met de antwoorden in het onderzoeksbestand. Die situatie is in de tekst niet aan de orde.

[55] Working Party 2010.

[56] Een andere situatie is ten behoeve van de eigenlijke patiëntenzorg zelf waarbij het elektronisch dossier bij een zogenaamde ASP is ondergebracht (vergelijk CBP 2009b). Indien een bewerker wordt ingeschakeld ten behoeve van de aanschrijven van (ex-) patiënten in het kader van de kwaliteitsborging van de betrokken zorginstelling, ligt dit naar mijn mening ook anders.

[57] Dit onderwerp wordt behandeld in de discussie over de herziening van de Gedragscode Gezondheidsonderzoek 'Oude' gegevens zijn nl. van groot belang om op de lange termijn veiligheid van nu gebruikte behandelingen te kunnen onderzoeken.

[58] Zie o.a. Herk-Sukel 2010 en de daar aangehaalde literatuur.

[59] Deze conclusie is gebaseerd op mededelingen van diverse onderzoekers.

[60] De heer C. Verhage, directeur van ZorgTTP, stelde in een reactie op een eerdere versie dat via een TTP kostenefficiënt wordt gewerkt. Indien behandelaar en onderzoeker bij elke verstrekking van gepseudonimiseerde onderzoeksgegevens zelf de codering zouden moeten ontwikkelen en de

audit zouden moeten organiseren, zouden de kosten aanzienlijk hoger zijn.

[61] Zie noot 26 voor de uitleg van het begrip aggregatieniveau in deze context.

[62] In die zin dat persoonsgegevens terechtkomen bij partijen die niet over deze gegevens zouden mogen beschikken.

[63] Bij zo'n situatie moet men denken aan het bekende voorbeeld van de arrestatie van een bekende voetballer in verband met een zedendelict. Het bleek dat dit dossier in het politieregister werd benaderd door talloze andere functionarissen, die niets met de zaak te maken hadden.

[64] Terwijl ik vele gesprekken heb gevoerd met onderzoekers om de groepsnormen en risico's op doorbreking van deze normen boven tafel te halen.

[65] In de zin van Nissenbaum 2010.

[66] Dit laatste wordt wel eens vergeten, waardoor paradoxaal genoeg die grote database beter is beveiligd dan de lijst met enkele patiëntgegevens.

ENGLISH PART

9. AN ALTERNATIVE TO THE VAIN QUEST FOR ANONYMITY

9.1 Introduction

This chapter is meant to propose a practical solution to the problem that not directly identifiable data in the research domain are often to be considered personal data by the standards of DPA's, even though researchers will consider them anonymous data.
The set up of this part is as follows. First I will very briefly discuss Directive 95/46/EC in relation to medical research with data. Then I will discuss to what problem(s) the TTP refers. I will show that making data even more 'anonymous' can not be the answer. The solution comes after that and discussion of remaining problems according to existing law.
I will finish this part by arguing that what is proposed here will give a temporary solution for some of the present problems but is not very satisfactory in the longer run, or even at present from a more meta-juridical or ethical point of view. That short discussion is the 'overture' to chapter 10.

9.2 Directive 95/46/EC and health research

9.2.1 In general

Though seemingly concise, Directive 95/46/EC is a complicated piece of legislation. As a Directive it had to be implemented in the binding law of the member states and did not as such have immediate direct effect(Lenaerts, 2011, 22-076 etc.) On several occasions it has been noted that the implementation has been very diverse in the member states regarding medical research with data (Privireal 2005, van Veen 2006, Verschuuren 2008).

This various implementation then refers to exemptions to the principle of explicit consent which is the basic principle to process sensitive personal data, such as patient data (art.8.2a).

These exemptions are the following:

- Paragraph 3 of art. 8 of the data protection directive allows for exemptions on the consent principle for amongst other things reasons of 'preventive medicine'. The data must then be processed by a health professional with a professional secrecy or another person with an equivalent obligation of secrecy. Though the Privireal (2005) project did not exclude that health research could fall under this paragraph, the art. 29 Working Party in its Opinion on EHCR's (Working P 2007b) considered health research not to be preventive medicine. The article should be applied restrictively, applicable to health care providers only.
- Paragraph 4 of art. 8 mentions exemption from the consent principle for reasons of substantial public interest, subject to the provision of suitable safeguards.

The Directive has further references to research.

- Art. 6 1.b states one of the basic 'privacy principles', namely that data should be collected for specified, explicit and legitimate purposes and not be processed in a way incompatible with these purposes. It then continues that 'further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible, provided that appropriate safeguards, as decided by the member states, are met'.
- Paragraphs 2 of articles 11 and 13 respectively provide for exemptions on the duty to notify the data subject about the processing of his data or to give him right of access to those for, amongst other things, purposes of scientific research, however only if certain conditions are met. In general it should be 'impossible or would involve a disproportionate effort'. Again member states should provide adequate safeguards.

9.2.2 Anonymous data

If data are anonymous they fall outside the scope of the Directive. See the summary of the Dutch part numbers 6 and 7 as to when data may be considered anonymous, based on the Opinion on the concept of personal data by the art. 29 Data Protection Working Party (Working Party 2007a). Yet, before they are anonymous, they have been patient data and, amongst others, Privireal (2005) questioned whether the anonymisation of patient data was not by itself a step which is covered by art 8. Connected to this Privireal mentions the 'broader conception' of privacy by which not all data which are relating to someone are covered but by which one should also be able to influence for which purposes data will be handled once they are anonymous.[67]

I will discuss the broader conception of privacy in the next chapter of this Report. Here it suffices to note that all member states recognise that patient data can be used for health research once they have been rendered anonymous. It is the starting point of the question of the Dutch DPA about the use of TTP, leading to this Report. Germany is a stronger case in point. It has probably the least exceptions to the explicit consent principle for the use of patient data for health research.[68] Instead Germany has strongly supported anonymisation procedures (Kühn 2004). In some instances, such as for cancer registries in some of the Länder, patients can opt-out based on generally available information, in other cases, as with the 'Forschungsnetze' there is no such opt-out procedure (Reng 2003). Yet, a broader conception of privacy as 'informational self-determination' has been recognised by the German Federal Court (BVerG 65,1 15 December 1983, in particular at CII1a)

From the text of the Directive the use of anonymous patient data for research without consent can be explained as follows. Art. 6, which states the basic privacy principles, also applies to the sensitive data described in art. 8. Hence also the fact that the further processing for historical, statistical or historical purposes is not incompatible with specific and legitimate purposes for which patient data were processed in the first place. Yet, if this further processing wouldn't render the data anonymous, article 8 would still be applicable to the dataset which is then created. But not if the data have been made anonymous. They fall outside the scope of the Directive, at least for the consent principles embedded in it. The historical, statistical or scientific purposes and the 'suitable safeguards' would still apply, as that is the condition under which they have been further processed and been rendered anonymous. The Dutch implementation of the Directive states that in that case they should *only* be processed for such purposes (art. 9.3 Wbp). This also shows that sensitive data, once they have been rendered anonymous, are not 'free'. E.g. a hospital could not 'sell' them to a pharmaceutical company for the commercial analysis by the latter of which drugs are usually used.

9.2.3 Controller, processor and the data chain

The Directive uses two other concepts which are important in this respect, being 'controller' and 'processor'. Controller is – in short - the entity which alone or jointly with others determines the purposes and means of processing personal data (art. 2.d). Processor is the entity which processes personal data on behalf of the controller (art. 2.e). Using a processor by a controller requires a legal contract between the two. The controller must ensure that the processor will adhere as well to all data security norms applicable to the controller (art. 17 section 2 and 3).
The law of the member state where the controller is based applies even if the processor is based in another country. The controller is accountable for the justified processing of the data according to applicable national law.

Long data chains and joint processing of data have made it at times more difficult to ascertain who is the controller or controllers of the data, and which entity or entities are hence mere processor(s). Determining the controller obviously comes first. The controller is accountable to the data subject.

In its Opinion on both concepts the art. 29 Data Protection Working Party emphasized that several elements should be taken into account to determine which entity is the controller (Working Party 2010). The Working Party chooses a functional approach. It comes down to who determines or can

determine for which purposes the data are to be used. The controller may have some discretionary role in determining how that is done, within the terms of the contractual agreement.

If two research entities would combine personal data to work with those jointly for a new research project, they together will be the new controllers. If, however, several research entities would deposit data in a repository held by a third party while each research entity still being in charge regarding what can be done with the data which this research entity has deposited, each research entity would remain the controller of his segment of the data in that repository. This would require as many 'controller-processor' agreements with that third party as there are research entities depositing data.

In the data chain for research it should always be clear who is the controller, possibly the processor, or the new controller. While on the internet the data subject might not know who is the controller and certainly not which are the processors, the data chain for research can be a much more closed one. From health care providers to research entities or an intermediary database such as a cancer registry and from there to research entities, to possibly new databases for joint research efforts. These concepts should be used correspondingly if, as will be forwarded in the next paragraph, in principle only pseudomised anonymous data are used in the data chain for research. It must be clear who is accountable for a certain dataset. Only then can it also be ascertained that data will be used according to the purposes for which the data were provided and for example will not be illegitimately re-identified.

9.3 The quest for anonymous data

If anonymous data would always be sufficiently suited for research, this Report would probably not have been written and it certainly would have been very much shorter. However, as has been argued by several especially epidemiological researchers in the course of this inquiry, anonymous data are often too general for more detailed analyses into cause and effect of onset of disease, variables which might have contributed to it and variables which might have contributed to the outcomes of treatment. By the end of 2011 the European 6th Framework Program project EUROCOURSE[69] intends to issue a detailed report on how the quest for anonymity and PET in-between, hampers especially cancer research.

It should be mentioned that the quest for anonymity already goes quite far at this moment. If the research exemptions do not apply, researchers will not get data if there only three rare cases in a given region (either a large rural region or a small densely populated town). I have met researchers who were told that they could possibly visit all the local cafés or read the obituaries in all the local papers to re-identify such cases. Hence they get the data on a higher aggregation level, assuring anonymity from this presumption of bad faith but often making them much less useful for their research. To mention just one example: it does make a difference if one gets all the data about survival of a hip replacement after an accident of all elderly people over 80 years or all those over 90 years (and with all the other data, such as interval between accident and time of operation, similar of course). This will even be more so when 'truly anonymous data' becomes a smaller and smaller category. The proliferation of personal information on the internet makes seemingly anonymous datasets identifiable by linking those datasets with public data found on the internet (Ohm 2010, Korff 2010). Ohm shows by the way that one would have to use rather sophisticated statistical tools for this linking. Yet it can be done and the presumption in the European debate is to a large extent that if it *can* be done, it *will* be done. So also large anonymous datasets , even when sufficiently aggregated according to present standards, might not be considered anonymous anymore. The traditional solution would then be further aggregation. But that would make those data even less suitable for research. To continue on this 'path to anonymity', as the Dutch DPA as once called the use of PET (Registratiekamer 2000), will lead medical research into a data desert.

9.4 A different approach, security within the research domain

The Opinion on the concept of patient data expressed the doubt of European regulatory authorities that pseudomymisation procedures in the datachain would be kept sufficiently secret (Working Party 2007). Korff (2010) argued that the Working Party didn't go far enough and had not sufficiently dealt with the present risk of re-identification once the data are at the end of the chain, in this case the research domain.
Yet, how realistic is it that researchers will try to re-identify patients or persons from the population at large? Do we really need 'regulation by technology' (Bronsword 2008) here or can we rely on the moral commitment of researchers, more than has been done in the past, that they are after patterns and not persons? A mix of both is proposed.
Researchers have no practical interest in doing that for their core business, health research, if they can get sufficient data (pseudonymised) from the data sources. They have no legitimate interest in re-identification outside that. There are no reported cases of such re-identification by a researcher and in my many discussions with researchers I did not hear even anecdotal evidence about such behaviour.

What has not happened yet, cannot be a sole reason for not preventing such an abstract possibility for the future. It should be possible on the other hand to establish such procedures and technical measures within the research domain that the processing of possibly indirectly identifiable data there can be considered sufficiently secure, both against illegal access to those data by third parties and to illegal processing, contrary to the purposes for which they were submitted, by researchers. Hence, the quest would not anymore be for anonymity *before* entering data into the research domain but for the safety and the integrity of the processing of those data *within* the research domain, applied to not directly identifiable data. Though much of what follows would also apply to directly identifiable data, in that case the situation is obviously different. Those data cannot be submitted as if they were anonymous. Such submission must be based on either explicit consent or on one of the research exemptions.

The discussion is then first of all about these procedures and technical measures. The outcome: if those are implemented, the data should be considered anonymous in the research domain. In terms of Recital 26 of Directive 95/46/EC, in that case there is no reasonable likelihood that means will be used to identify the data subject.

I will discuss those procedures and technical measures in Appendix 1. Here it should already be mentioned that technical means cannot completely with 100% certainty pre-empt illegitimate processing by researchers. Procedural measures should also count. Together they can ensure that any illegitimate use of data is extremely unlikely and if it does happen, this would come to light soon enough and adequate measures against the wrongdoer will be taken.

In this context I should forward the following. I cannot imagine any social 'practice' (governmental, business or private) which is more dependent on data security and legitimate processing than that of (epidemiological) research. Such researchers are dependent on the trust data sources have in researchers. When discs with social security data are lost, people will still apply for social security. When a commercial business has been sloppy with personal data, it might be fined and there will be bad publicity, but most consumers will still buy the product if it is sufficiently competitive. Millions of people submit their personal data on Facebook or send intimate mails via g-mail, in spite of the privacy issues connected to those services. However, if researchers are sloppy with the data submitted to them, that will be much more than just – after the Dutch expression - a 'ripple in the pond'. The dataflow for this particular project will stop. It will most probably stop or temporarily halt the dataflow to other projects. It will have a huge negative impact on the career of the researcher involved, might even be the end of it, and will most probably also impact the careers of his colleagues.
These more 'practical' considerations come in addition to the inherent norms of researchers. Many

are epidemiologists and hence doctors obliged to an oath of confidentiality. All are members of 'learned societies' and subjected to codes of conduct and the scrutiny of their peers. Epidemiological research is – to say the least - not the most profitable career path in medicine and health research in general is not the most profitable career path for a new student. It is usually chosen to contribute to the public good.

These researchers can be seen as 'professionals' as Freidson saw them (Freidson 2001). In his last major work Freidson forwarded 'professionalism' as the 'third logic' between the market and bureaucracy. Though critical about the medical profession in his earlier work, he was concerned about the erosion of professionalism as a cornerstone of inherent values, which do not only protect the professionals but also their clients, in-between those two other seemingly dominant forces of present society. Professionals in health research might not fit completely with Freidson's 'ideal type' of professionalism[70], yet they come very near. And they do have clients, even if they usually do not interact with them directly. Researchers see those who can profit from their research (future patients or those who will not become patients because of the outcomes of research) as their clients. Which also leaves researchers rather vulnerable as such clients can hardly form a pressure group and do not pay them directly. Elsewhere I have forwarded proposals how the interaction with their "clients" could be reinforced (van Veen 2008).
The erosion of professionalism seems even more poignant today and the kind of proposals made here fit in with Freidson's approach. Also in the sense that professional values cannot be taken for granted but must be made explicit and transparent and defended through such practices. And in this technological age, must be corroborated with appropriate technical measures.

Recognition of research organisations as 'trusted entities' is not new. I made a plea for that in 2008 (van Veen 2008) and Ohm (2010) was more explicit on it, in the light of the challenges of new re-identification methods due to the proliferation of patient data on the internet.

9.5 Additional issues

The above means that data in the research domain can be considered anonymous data – if certain criteria have been met - , yet the same data would not be considered anonymous data in the hands of another entity, such as a direct marketing company. Does this mean they *are* anonymous data in the research domain? This raises the question whether Directive 95/46/EC allows for such a contextual approach. In my opinion it does. I will come back to that in chapter 11.
At this stage there is already an interesting paradox if such data would not be considered anonymous data in the research domain. If they would be considered patient data, also the provisions about notification etc. of the data subjects would apply. These can be mitigated by the research exemptions as implemented in national law. As seen, the Directive allows for non notification in the context of research if this would involve a disproportionate effort. The paradox is that in order to assess whether it would be disproportionate, the researcher must first of all do what he shouldn't do, namely re-identify the persons in those datasets at the research domain. Unless one would consider this re-identification by itself disproportionate (as well as being illegitimate). But in that case one may rightly wonder why those data are not considered anonymous in the first place. This paradox seems to be overlooked in the literature, such as in Curren 2010.

This holds even more true for alternative ways for transparency. There should be transparency about what research with data is going on and a world can be won here. However, that is general transparency to the public at large. If a patient were to inquire about data regarding him or herself in a particular research database, submitting his or her name, address, birth data, the researcher would not be able to find that patient. Patients in the research database are too much pseudonymised for that. It might be possible through the data source, though whether that is feasible will very much depend on the layers of pseudonymisation which have taken place since then. Data might arrive two way coded in a cancer registry (assuming that such a cancer registry is allowed according to national legislation) and might end up unlinked in a following research database. Yet, these data

might still be regarded as personal data because of their aggregation level according to the very strict standards which are challenged here.

The next issue is more problematic. The Directive does not cover anonymous data. The processing of personal data are subject to various requirements. The distinction between personal data and anonymous data has always known a grey area and present developments have made the distinction even more problematic. The proposal made here, in connection with Appendix 1, will solve many of the present complexities, but also shows that the all (requirements for patient data) or nothing (the absence of those for anonymous data *if* the are used for research) approach becomes rather arbitrary. That is even more so in the light of the 'broader' conception of privacy referred to in section 9.2.

The consequences of this conclusion can only be established in the context of a broader discussion on the conditions under which patient data may be used for research. That will be done in the next chapter. It will follow from that chapter that patient data should be used for health research for the public good, seen as research from which other patients might profit or which will protect the public from becoming a patient. Such research can even involve personal data without consent if that cannot be avoided. Yet, as will be forwarded in that chapter , an opt-out option should always be offered. In the concluding chapter I will rehearse this option together with other challenges in concrete proposals both for the existing Directive and coming update of the Directive.

10. THE DEFENCE OF RESEARCH WITH PATIENT DATA

10.1 Introduction

Medical law has greatly contributed to neutralise the uneven power balance between doctor and patient. More patient empowerment is to be achieved by considering the patient a consumer of health care services. The patient will not only be able to choose about his treatment within the consulting room, the principle of informed consent, but can also make an informed choice about – in short – which consulting room to visit. The recent Directive 2011/24/EU on cross-border health care can be seen as an expression of this. Insofar as this will make the health care system more responsive and authorities more accountable for the choices made, this is can only be applauded. Yet, this approach also has inherent limits, following from the essence of national European solidarity based health care systems (Newdick 2011). Additionally there are several major aspects where patients differ from other consumers. Not only that his or her life or well-being is at stake and that health care is embedded in an intricate web of regulations concerning rights of patients and quality of care.

The following are relevant for the present discussion:

- European health care systems are based on principles of solidarity and equal access, together with quality assurance and a system of cost-containment to assure long term sustainability of the system (Commission 2004). Entitlements to health care are based on statutory provisions. As such they are need-oriented and not demand-oriented.(Bilsen 2007). With the exception of co-payments schemes which are marginal considering the total costs of his health care (Tambor 2010) and interventions which are not covered by public schemes, such as in purely cosmetic surgery, the costs of a patient's care are borne by the collective, irrespective of whether he[71] will have contributed to that collective.
- The next element, very much connected to the former, which differentiates health care from any other market, is that the protection from health threats and the equitable access to health care is considered a human right (Den Exter 2008, see also the Charter of Fundamental Rights of the EU, art. 35).
- There is a third extremely relevant difference in the present context. Unlike any other realm of consumer law, the 'consumer' in health care is not only a purchaser of a product or service but the end product itself as well. His condition in terms of health and well-being is the outcome of the services he purchases and hence in that respect the patient also *is* the product instead of just having acquired the product or service. Medical (or nursing etc.) services are an intermediary phase to the desired result: better health (or if that cannot be achieved: the most possible wellbeing, comfort etc.) for the consumer, the end product.

These three basic characteristics of health care delivery also influence the nature of patient data. A doctor cannot treat and a nurse cannot give care without knowing very personal and intimate details about the patient. The salient point here is that many of these details were not known to anyone beforehand but are generated during the diagnostic process within health care, such as lab findings or findings from imaging techniques. And the response to treatment will generate new data about the particular patient, like the non-response or even allergic reaction to a certain drug which performs well with most other patients. Repeated administration of that drug to this patient should be avoided.

This is the *individual* function of patient data. It is hardly ever called as such, as the corollary, the *collective* function of patient data is often overlooked. The individual function and privacy protection necessary in that context (see e.g. ECHR 17-10-2008 I v. Finland) has got most emphasis in medical law and the collective function, if discussed, is seen as the exception. A clear example of this approach is the Working Paper of the art. 29 Working Party on electronic health care records (Working Party 2007b).

Here it is submitted that the collective function of patient data is just as important as the individual function. The collective function of patient data makes equitable health care systems work and sustainable in the long run and protects patients or the public in general from health threats. This collective function limits consumerism in health care, seen as, related to data, the patient being in control of which data go where. The collective function might even trump individual privacy if certain conditions are met.

10.2 Various applications of the collective function

10.2.1 Introduction

This collective function is not one unity but knows various applications. Though in all the necessity and proportionality of the use of patient data must be considered and PET must be applied as much as possible, the result of that assessment will differ for each of these applications.
In general the following can be distinguished, partially overlapping each other:

- Patient data for reimbursement schemes;
- Patient data for monitoring the health care system, allocation of resources through statutory schemes and the like;
- Patient data for surveillance and containment of infectious diseases or any other epidemic;
- Patient data for public health research and medical research in general.

This Report is about the last category and hence the emphasis in this chapter will be on that. However, to clarify that also that last category is not exceptional, I will give some examples of the others, most of all drawn from the Dutch debate.

10.2.2 Within the system of reimbursement and cost containment

The Netherlands has perhaps gone the farthest of all EU Member States in creating a market-oriented health care system within the boundaries of a social insurance system. One pays a direct premium to the insurer. The insured has the right to choose the health care insurer and change insurer every year. Apart from the services offered, insurers compete with the amount of the insurance premiums. That premium is for most income groups much less than what is paid through taxation. Those tax premiums are transferred to a health care fund. Through a 'risk equalisation scheme' that fund is divided among health care insurers. Health care insurers contract health care providers to deliver health care to their policyholders, and can demand that they adhere to certain standards to acquire the contract. The system is quite correctly considered to be a social insurance system in terms of exemptions for social insurance schemes in the non-life insurance Directives (Van Veen 2005). One cannot choose not be insured. Health care insurance is compulsory and health care insurers are obliged to offer the statutory scheme. The types of health care which fall under these statutory benefits are regulated. Insurers are obliged to contract this health care or compensate for the costs of the patient if the health services are delivered by a non-contracted health care provider. The largest part of the costs is still born by taxation, either through the abovementioned fund or through reimbursement of direct contribution for lower income groups from tax revenues. Health care costs within the compulsory system are considered to be collective costs and the total national budget for this health care is annually laid down by the government.[72]

Without going into the details of the Dutch system, it is obvious that many data need to be exchanged to make this system work. The most direct exchange of patient data is the one required to reimburse the health care provider by the health care insurer. The Health care insurance Act (hereinafter: HciA) stipulates that the health care provider is obliged to refer certain patient data to the health care insurer for that purpose.[73] These are rather general, based on 'treatment diagnosis groups', but still refer to the patient's condition. This led to an interesting dispute when short term psychiatric care was brought under the ambit of the HciA.[74] Psychiatrists claimed that the system jeopardised

the privacy of their patients and would be detrimental to the accessibility of psychiatric care. The discussion is still continuing. At first the governmental agency which regulates the reimbursement system, the Dutch Health Care Authority (NZa), stipulated that these objections were exaggerated and that those data were anyhow necessary for the proper functioning of the system. In his latest decision the administrative Court, where the discussion ended for the time being, was not convinced by the arguments of the Agency but neither were the claims from the psychiatrists fully conceded. The Court decided that the NZa had not been able to prove that the required amount of details in patient data were necessary and proportional to the ends of the reimbursement system. It did acknowledge, however, that some patient data were necessary to that end and hence the health care insurer would need to know something about the patient's psychiatric treatment in order to assess eligibility of the patient under the HciA and the rightfulness of the reimbursement to the health care provider. The agency was charged with finding a better balance.[75]

The NZa has the statutory duty to guard patient interests on the health care market, which can be divided into the insurance market and the health care providers' market. But it has also a duty to guard the proper functioning of the system as a whole and issue Directives about what prices may be charged for certain types of health care. Together with other agencies it can make inquiries into fraud and require access to all necessary files to that purpose[76]. These may include patient data insofar as that is necessary and proportional to a specific inquiry[77].

10.2.3 Monitoring the health care system and health care providers

To meet the basic informational needs of the Department of Health and the many agencies charged with the execution of the HciA, like the agency which administers the risk equalisation scheme between insurers, a complicated system has been set up where health care providers are required to send pseudonymised data[78] to a portal. From there this information is sent to these various recipients.[79] The pseudonymisation is performed by a Trusted Third Party. The pseudonymisation is reversible by one of the recipients. The Dutch Central Bureau of Statistics (Statistics Netherlands) has the statutory authority to receive personal data of all Dutch citizens[80] and is able to reverse the pseudonymisation and hence link the health information with other records available at Statistics Netherlands. Though the informational needs of the recipients are based on statutory provisions, the actual working of this information flow with the Trusted Party and the portal is based on a practical understanding between all parties involved, and its governance structure is not published officially, neither by the government, nor by the entity which administers the portal. In sharp contrast with the individual function, this collective function has hardly been discussed in parliament or in the legal literature.

Patient data are also processed for monitoring the functioning of separate health care providers. The Inspectorate for Health requests hospitals, nursing homes and the like to submit statistical data of their performance on certain indicators, such as the incidence of decubitis etc. These data are based on the patients' records and because of the complexity of some indicators, an underlying datawarehouse and business intelligence to make them available as 'indicator data'.
A consequence of (the presumption of) competition in the Dutch health care system is also that health care providers need to be transparent about their performance to prospective patients. Each year a set of 'indicators' is established for this purpose. The data are assembled by a health care provider in the same way as for the indicators for the Inspectorate for Health together with measurements of patient satisfaction. They must be submitted to an organisation "Visible Care"' which will then publish them on the internet.[81]
The scientific validity of these 'indicator data' has been challenged as they lack, amongst other things, to a large extent 'case control mix correction'. [82]

The Dutch system might be a bit more complicated than others because of its mix of market elements with solidarity based health care insurance. However, in all health care systems information is gathered about health care consumption and performance leading to policy decisions by the competent

bodies. In a 'Bismarckian' system the health care insurer will receive personal information about care consumption as well. Although this is done to assess the eligibility for health care benefits of the insured patient and to reimburse either the patient or the health care provider (or a combination thereof), this cannot be considered a solely individual function. It is a function which is closely connected to the general, collective aim of the health care system: ensuring equal access to care based on need, cost control and measuring the overall performance of the system.

10.2.4 Data and communicable diseases

There are many threats to health. Some of those are inflicted upon us by our environment, like exposure to toxic substances; some of those we might inflict upon ourselves, like smoking; some of those we carry inside, like the propensity to high blood pressure; and some of those we might transfer to another person, namely certain communicable pathogens.[83] In any developed health care system it is essential to know what pathogens are around and whether there is an increase in them or not. The health care system must be able to react to such increase. The reaction might take many forms, ranging from alerting physicians to uncommon disease characteristics which might be due to an uncommon pathogen to a new vaccination program. Based on – at present – article 168 TFEU the European Union a surveillance program has been instituted for communicable microbiological pathogens[84] thus reinforcing national surveillance programs[85]. Such surveillance programs are dependent on patient data, as they start from there. A patient presents himself with certain symptoms and then during the diagnostic process it is found that the cause of the patient's disease can affect others as well.[86]

There is a much sharper edge to patient data in the context of communicable diseases. For the containment of such diseases, sometimes measures must be taken by which patients, or carriers who might not be sick at all, must be found by name and address. This might take various forms which cannot be fully discussed here: from finding, through the voluntary cooperation of all involved, the so-called index patient who is the carrier of a disease that is relatively mild for most but which can be dangerous in special cases, such as measles,[83] to compulsory detainment of the carrier of a disease which might end up being a disastrous pandemic for many. The International Health Regulations 2005 of the World Health Organisation (WHO 2005) require all parties to that Treaty to have public health laws in place which can monitor highly contagious diseases, detect new unknown pathogens, and take compulsory measures for containment when strictly necessary. One can always discuss the balance which has been struck in those public health laws between the privacy of the individual and the protection of others (critical about the Dutch situation, Dute 2008). Yet, the fact that at a certain point privacy or autonomy of the individual must give way to the protection of others, and that identifiable patient data are needed for that purpose, is accepted by all.

10.3 Patient data for health research

10.3.1 Introduction

The last part of the previous section might seem an exceptional example of the collective function of patient data. Yet, there are many more examples outside the context of pathogens which are carried by our fellows. The list of such examples is nearly endless and includes, for example, salmonella in chickens, Lyme disease through tick bites which we did not consider a threat a decade ago and, going back in time, the risk of smoking or, probably the earliest example of epidemiological research, the correlation between water wells in certain neighbourhoods in London with cholera (Porter 1997, p. 412). All these risk factors have been found through patient data and combining them with environmental data. The same holds true for risk factors which are due to health care. In that case, the risks are found by combining outcome data of *the product itself*, the patient, with data from earlier treatments (and often combined with lifestyle data and environmental data to exclude confounders). Again, there are too many examples here to give even a brief summary. They range from the discovery of blood subtypes which has made blood transfusion immensely safer, to

the discovery of adverse reactions to many pharmaceuticals.

Advances in medicine and in health care services do not just drop from the sky. They are based on research. Though most would probably think of clinical research, such as clinical trials, probably the most important and first source for this research is (epidemiological) research through the use of patient data (Lowrance 2002, AMS 2006). This means a systematic looking back to what is actually happening with groups of patients. In order to explain – or find the beginning of an explanation in the form of a new hypothesis – why this is happening, these patient data usually must be complemented with additional data, such as environmental and/or lifestyle data, and should have sufficient statistical power to exclude confounders. This is a collective function of patient data. They are used for the public good. It is one of the foundations of health care as we know it and an essential element for further improvement. It cannot be seen as an exception to the use of patient data, as it is done by, amongst many, the Article 29 Working Party in its Opinion on EHCR's. (Working Party 2007b)

10.3.2 Objections : general overview

Even if this tenet was acknowledged, it is still possible to raise many objections. These objections range from pragmatic ones to more principled ones. I will discuss the more principled objections in the next section. The first more pragmatic objection will be that even if we need this type of observational research, this can be very well done with anonymous or pseudonymised[78] data.

In the previous chapters this illusionary idea has been discussed already. Monitoring the health care system can be done very well with anonymous data. But the best of both worlds, namely both complete privacy protection through only using anonymous or pseudonymised data and still being able to perform good observational research, does not exist. It is a completely different point, however, that datasets entered in the research domain will not become available in the public domain or that indirectly identifiable patients in the research domain will be illegally re-identified there.
The other objection is that this kind of research can very well be done with explicit or informed consent. That objection is discussed in the next section. It will be shown that consent does create problems in acquiring sufficient reliable data for health research. The fact that it does, is, however, by itself not a legitimate reason to abandon this principle in this context. That can only be so if we can decide that foregoing consent does not compromise the patient's autonomy in an unacceptable way. That issue will be discussed in paragraph 4 of this chapter.

10.4 Problems with consent for using data for health research

10.4.1 Terminology

First a brief terminological explanation. Directive 95/46/EC uses the term 'explicit consent' for processing sensitive data outside the exceptions to processing them without consent. With explicit consent is meant general information about what will be done with the data, who is the processor of the data and about possible third parties to which the data might be transferred.
Medical law uses the term 'informed consent'. That refers to a much higher threshold for consenting. The patient should know what *specifically* will happen when he consents, the possible side effects and the alternatives. Medical (case) law often also differentiates between the information required for purely cosmetic procedures on the one hand and procedures which are doubtlessly necessary to save the patient's life on the other. The threshold for the necessary information concerning possible side effects is much higher in the first instance than in the latter. The highest threshold for informed consent is used in consenting to a medical trial. These are not only pharmaceutical trials as covered by Directive 2001/20/EC but any procedure which will deviate from the standard of care in order to acquire scientific data. In such a case the information should be on paper, it should be specific about the scientific aims, the advantages (if there are any) and possible disadvantages of participating,

there should be waiting period and the patient must sign his agreement to participate. The later is in my opinion not primarily to prove that the patient had consented but make the possible participant – once more - aware of the substance of his decision.

Hence there is a substantial difference between explicit consent and informed consent, at least in the discourse of medical law. It is interesting that the European Commission mentions in its document on the proposed policy for the renewal of the data protection 'informed consent' instead of 'explicit consent'. (Commission 2010). With 'informed' instead of 'explicit' the Commission may have wanted to indicate a higher threshold for consent to processing personal data than at present in the Directive. That could be an ominous sign for lessening the administrative burdens which according to the Commission document mentioned the renewal of the Directive should achieve as well. Or there will two conceptions of informed consent, depending on the context, namely one for privacy protection and one for medical procedures.

In a discourse which addresses both medical law and privacy protection two different conceptions of the same term is confusing. I will use the two different terms. Explicit consent is used for general information about data processing, informed consent for much more specific information.

10.4.2 Overview

Both explicit consent and informed consent raise issues which are problematic if one considers that data of all patients as end products of health care can be a valuable input for research:

- It is usually being asked at a moment which is very inconvenient for the patient, especially when the treatment options for a life threatening disease must be discussed. The dialogue between doctor (or nurse) and patient is meant to get a complete understanding of what is at stake and an informed choice from the patient as a result (or several, step by step as the diagnostic process, treatment and sometimes the disease progress). There is usually no mental headroom on either side to discuss the secondary use of patient data. There is usually no time either. Other patients in distress are waiting. I will come back to that soon;
- This can be different in situations such as:
    - Patients with certain chronic diseases who visit their doctor or nurse- practitioner regularly and where disease management is fairly stable;
    - Parents of children with a serious disease. Though they are existentially worried of course, they are affected in a different way. They are not seriously ill themselves with all the fatigue and mood changes that brings with it. While the patient is being treated they can, as the patient's guardian, discuss the treatment. Also because of the dedication of paediatricians, there is more time to discuss additional issues.
- To reach true explicit consent will take time and is costly. In a study on consent for 'further use' of residual tissue it was found that most patients did not remember having given consent, even though they had been informed and had even signed a simple form (Mancini 2010);
- This consent can never be informed consent. The data protection in the chain can usually be explained in general terms, and the absence of any consequences for the present treatment, but not for which health research exactly the data will be used as the specific questions are not known beforehand;
- That would be different if one asked all patients retroactively for each and every protocol. Except for patients with chronic diseases, that can only be done after the treatment has ended. Patients will have to be re-contacted, insofar as they are still alive, and often more than once, as, especially for cancer, data are used in many research projects. This obviously creates huge costs and hindrances of its own;
- Explicit or informed consent has been shown to create bias in the data, often detrimental to already disadvantaged social economic groups (Jacobsen 1999; Woolf 2000; Tu 2004, Al-Shahi 2005; Jousilahti 2005). Cancer registries have ended when consent was introduced (Ingelfinger 2004 and more examples in the coming report of the EUROCOURSE group).

10.4.3 Taking informed consent seriously

It must be stressed that the above problems with explicit consent or informed consent for research with data (or with residual tissue for that matter) are mentioned against a background of taking informed consent in health care very seriously. A (serious) disease compromises the patient's autonomy. Health care is meant to restore the patient's autonomy as much as possible. Without informed consent about diagnostic and therapeutic procedures, that can certainly not be achieved. This should be embedded in interpersonal skills to communicate with respect and compassion, especially when 'bad news' is the message (Poulson 1998, Ten Haaft 2010) or when the options are very limited (The 2010).
The patient's vulnerability during this process shouldn't be underestimated. A serious disease often leads to mood changes, either because of obvious psychological reasons but often also as the disease or the medicines lead to metabolic changes in the patient. The diagnostic and treatment process is often incremental, constantly re-evaluated and adjusted in the light of new challenges for doctor and patient alike. Even when the patient will be cured or a stable disease management can be achieved, there are hardly 'quiet moments' as I once heard a civil servant optimistically describe the opportunity for informed consent for 'further use' of residual tissue.

One should not generalise too much here. Patient organisations have been formed. Especially those afflicted by genetic disorders, or their parents, have been very effective in self organisation (see also examples in Rose 2006) and as a drive for change[87], often also, by the way, with a plea for less restrictive regulation concerning health research (Van der Valk 2010).

Though as shown in Ten Haaft 2010, even doctors are usually absorbed by their own disease when that is serious, and even often need guidance. Today patients can find valuable resources on the internet, etc. Though what patients find on the internet will often be at odds with the evidence-based medicine which the physician must adhere to (Broeren 2011). Obviously I am less optimistic than Veach about patients healing themselves (Veach 2009). See also Agledahl 2011 for a much more pessimistic view. That might be the other extreme though, not only considering real patients asking for help from their doctor but also taking health care as it is too much for granted. I do agree with the importance of Veach's arguments of a normative bias in many clinical practice guidelines, protocols and the like.
Regaining the patient's autonomy in the face of disease should remain the highest goal. At the same time we should not see patients as abstractions but as real people, vulnerable, often older and most of them not at the educational level of the participants of the bioethical and legal debate. There is sufficient empirical evidence about the 'therapeutic misconception' of patients entering clinical trials (Van Leeuwen 2008) or of parents deciding for their children in such cases (Eder 2007, Fisher 2011). In other words, under circumstances when the informed consent process was meant to achieve the highest level of understanding of the choices made. That might be due to the fact that doctors, trial nurses and ethical review boards didn't do what they should have done, namely assuring well informed choice. But it might also at least partially be due to the fact that patients are not always what we expect them to be. If ethicists and lawyers keep seeing patients as abstractions which have to fit a normative framework as 'autonomous agents' navigating through health care on their own compass , we could end with the contrary of what we hope to achieve, and might become – after Macklin – 'enemies of patients' too (Macklin 1993).

10.4.4 Consent for using patient data for health research versus informed consent for medical interventions.

The above means first of all that it might not even be appropriate, given the challenges for many patients to regain control over their life and well-being, to confront them with questions about whether data which were collected for their individual function, may be used for the collective function at a later stage. That would be different if choices about which data may be used for which research would be an essential element of regaining the patient's autonomy through health

care (either in the consulting room or adjunct to it).
But that is not so. As will be argued later, health research with data will not in any way affect the present patient's treatment or well being or social status or whatever at a later stage, if certain conditions are met.

It does affect the patient's autonomy in a different conception of autonomy, namely that of informational self-determination. There has been a confusion of those two issues, because it concerns the same data and - seemingly - the same moment in time. The patient is there anyway so it seems logical that questions about choices about informational self-determination can be asked as well. As argued, that might be a very inappropriate moment though. Another source of confusion is in the terminology. The patient's autonomy in the face of disease is certainly the Dutch discourse often translated as the right to self-determination (Legemaate 2004).[88] Informational self-determination then seems to follow from it. Yet, this is different concept.
A more genuine reason to confound the two issues is that the patient must be confident that all those data which become available, either self reported or through diagnostic procedures, will not be used against him. That is the foundation of medical secrecy. The mixture of these reasons have led to keeping researchers at arms length from patient data.

Now that the two issues have been distinguished analytically, it still must be established whether the patient should or should not have this right to informational self-determination and whether using data for health research indeed does not negatively affect the privacy context of health care, his present treatment and later life. Those questions are interrelated. They will be discussed in the remainder of this chapter.

10. 5 A critique of fundamental challenges to data for health research (without explicit consent)

10.5.1 Beyleveld on micromanagement of data

A strong principled argument against using personal data for research without informed consent is found in Beyleveld's essay on using patient data for research (Beyleveld 2007). First, he states that even anonymous data would fall within the ambit of Article 8 of the European Convention of Human Rights and of the European Data Protection Directive. That argument is obviously wrong as has been argued in the previous chapter of this Report.

More interesting is a hypothetical case brought up in this essay of a patient with a gynaecological problem. I will expand the example a bit. Let us call her 'our conservative aunt'. She has strong religious opinions about what is right and wrong. She is opposed to chemical contraceptives and probably does not endorse *in vitro* fertilisation (IVF) and many more modern applications of gynaecological treatments. She might allow the data from her treatment to be used for the improvement of a very similar condition as her own, but would strongly oppose any other use which might improve treatment for practices to which she opposes or which can be the consequence of such practices. To Beyeleveld that is the main ethical basis for prohibiting the use of her data for such purposes, even anonymously.

Yet, the real world of health care is immensely more complex. We know little about our conservative aunt (as these are usually not the kind of issues we discuss at family meetings) but the chances are high that her treatment is more expensive than her contribution to the health care system. The funds from which also her treatment is paid might come from sources which she would not endorse at all, like the taxation of (employees of) pharmaceutical companies which market contraceptives. Her doctor will have gained experience over the years, also by treating patients with dissimilar conditions, but experience which may be needed for this particular treatment as well. If a surgical procedure was necessary, the surgeon would use his expertise with operations on ovarian cancer. A major contribution to the onset of this cancer is infection with the human papilloma virus (HPV). HPV is sexually transmitted[89]. But also with more mundane pharmaceutical treatment, the treatment

will be based on experience gained from realms beyond the present application. Most probably it might have started with the use of animals for research, which our conservative aunt, knowing her at least that much, is likely to oppose as well. Medical and pharmaceutical knowledge is an intricate web of knowledge, with immense cross-fertilisation between the various fields of expertise and research results.

Hence, if our conservative aunt profits from all that previously acquired expertise for her treatment, yet opposes any further use of her data except in a very limited realm, the least we can say is that she is not very consistent in her views if she would know about this immense cross-fertilisation. We might even call her a 'free rider' then with her refusal to allow her patient data to be used for further improvement of health care in general. But most probably she does not know, just as most of the public is unaware of this. And anyhow, health care has its own sphere of justice (Walzer 1983) and compassion should count more than simple consequential reasoning.[90]
Of course she should get the best available treatment and we probably should not tell her how that has been made possible. We should even try to accommodate her views about future use of her data, as with an opt-out system. I will come back to that later. But Beyleveld's example cannot be the ethical basis for the informed consent principle of all use, in whatever form, of her patient data. On the contrary, it shows the opposite. Beyleveld's approach would lead to, as has been described elsewhere, the 'tragedy of the anti-commons' (Bovenberg 2006, referring to ownership of residual tissue).[91]

10.5.2 This is biomedical research to which the principle of informed consent should apply

The origin of all international Conventions[92] and non-binding Recommendations[93] concerning the protection of participants in research goes back to the atrocities committed in the second World War, though there have been other very serious infringements on the protection of persons involved in research outside and before that context[94]. The common dominator of this kind of 'research' was that it was 'interventional'. The conditions of the person concerned were changed. In certain cases nothing was done, so called to follow the "natural progression" of a disease. In those cases there was, without the patient's consent, a deviation from the standard of care to which any doctor should be committed and every patient may expect from his doctor. That amounts to an intervention.

However, nothing of that kind happens in the case of the use of patient data for health research. It is observational research, which is performed in the background and often after the patient has been treated. I have discussed this fundamental difference elsewhere (Van Veen 2008, Riegman 2011) and especially recent ethical literature seems to recognise this difference and its consequences for the consent issue as well (O'Neill 2004,Knoppers 2005, Hanson 2009, Sutrop 2011 with further references, Campbell 2011).

10.6 Towards a legal presumption that patient data may be used for health research

10.6.1 Introduction

Yet, all of this does not specifically legitimise the use of patient data for health research. It merely shows that some of the main arguments against it are not very convincing

Ethicist John Harris once pleaded for a 'duty to participate in research' (Harris 2005). The basic moral rationale was that it can save human lives and, when possible, we are obliged to do so. The refutal of this argument was in essence that such a duty is at best a 'supererogatory' moral 'obligation', one which comes with many others but which one cannot be obliged to do (Shapshay 2007).I agree with that and can easily make a bolder statement. A legal duty to enter into clinical trials would be outrageous.
But we have a completely different situation here. Patients are not required to do anything. There

is no risk or burden, except that their data will be used. I will discuss the 'risk' of such data use later in this chapter. It suffices here that this risk, if existing at all, is of a very different nature than participation in trials as meant in Harris' article. Trials always carry a certain risk to health, even to life. Additionally, using data for health research does not even require an active contribution of the patient. The discussion is about *allowing* such data use which happens at the background of health care.

Again, this does not automatically lead to a binding obligation to allow such use. It does show that arguments that this would be a mere 'supererogatory' moral 'obligation' are much weaker.
Under those circumstances it comes down to making a balanced choice which can be defended on grounds of (social) justice. It will be a supererogatory duty to contribute to charity above which we pay as taxes. Paying those taxes is required. In Europe we all contribute to the health care system via taxes or a very similar retributive system. The richer contribute to the poorer, the healthy to the sick or handicapped to achieve equal access for all. Our 'autonomy' to decide about our income has been seriously curtailed in that respect. Of course, we all acquired that income against the background of a well organised society and couldn't have acquired it without that fabric of institutions, including equitable health care.
The same applies correspondingly to data for health research. The treatment of a patient and the data which were noted about this treatment, wouldn't be possible without all the background knowledge and expertise acquired by the treatment of all previous patients and research on those data. Why shouldn't a patient be required to contribute by the use of those new data for health research to further improvement to health care for others, or for himself if those results would still come in time?

10.6.2 Inventory of counter arguments

The arguments against this reasoning are manifold. The presumption here is that the reader supports the solidarity based approach to health care in Europe. If not, we don't discuss on common ground. Within that broad context of social justice it can be argued that someone's (patient) data is different from someone's money. Obviously there is a difference here. Income can be re-acquired and most of all do, each and every year. Privacy lost, is lost forever. Yet, this also depends on the conception of privacy in the context of health care, where researchers are by many seen as alien to that context. I don't agree and will come back to that soon.
The next counter argument is that those data might also be used against the patient, either by privacy breaches by which they will become known to third parties or as the result of the use of fully anonymous outcomes of research. The first aspect of this counter argument has been discussed already. The second aspect refers to so called 'group privacy'. This will be discussed in the next paragraph. The conclusion of both discussions is that they are not sufficiently valid counter arguments.

A third argument type of arguments can be that this availability of patient data for research might be counterproductive to equal access to health care. Patients might become afraid to visit a doctor. Medical secrecy is meant to guarantee equal access, irrespective of what data the patient might reveal during his visit (van Veen 2003) To a large extent such an argument would be based upon (and perhaps even fuel) a wrong perception of what data are used for research and how they are used for research. This data use for research does not endanger the patient's secrets to parties who should not interfere in the context of care. But it cannot be excluded that certain patients might not trust it, are with their back to society already or whatever reason they might have, and that the idea that data will be used for research can create a barrier to consulting a doctor. In that case the use of data for research without consent would be counterproductive to what it means to achieve, better health care for all. Hence the opt-out option which I will propose at the end of this chapter and to which I will come back in the concluding chapter.

10.6.3 This is not about the patient's autonomy within health care

The main counter argument hasn't be fully discussed yet. That is that we deny the patients' 'autonomy' to decide about what data will be used for what purpose.
That isn't denied in the individual function. At least according to Dutch medical law, the patient could decide that another health care provider is not allowed to see his data from the earlier diagnostic process even if those data would be from an objective medical view – or that of the average reasonable patient - necessary for the next phase in his treatment. In this case the exercise of autonomy to decide which data go where can be seen in the context of regaining the patient's autonomy in the face of disease. Some might say that by doing so, the patient does not regain his autonomy at all,[95] but that is not as this (exceptional) patient sees it.

But here we have an application of the collective function. Contrary to the individual function, the decision which data go where is not necessary to steer his treatment as the patient wants it. As argued, this is not about autonomy or self-determination in the face of disease, but about informational self-determination. In this case the patient will not bear the consequences of the exercise of this autonomy, but others will. We can curtail autonomy in the sense of 'informational self-determination' in many realms where private and public interests meet. The cited decision of the German supreme Court (paragraph 9.2) recognised that as well. Against the background of the solidarity based health care system it is legitimate to curtail it here, in order to improve health care for all. As mentioned, the contrary could lead to the tragedy of the 'anti-commons'.
Which certainly does not mean that everything goes in this context. As mentioned, there should be an opt-out function (more about that the last paragraph of this chapter). Necessity, proportionality and data security should be assured. I will come back to that in the last chapter.

First I need to discuss two aspects which were promised earlier. The first is the suggestion that using data for health research is much less contrary to the privacy context of care than is often assumed. The second is that the results of research do not carry a danger to groups of patients or others.

10.6.4 Patient data for health research as contextual integrity

In an outstanding book on privacy Nissenbaum has sketched various conceptions of privacy, most of all taken from American literature (Nissenbaum 2010). She pleads for an approach to privacy as 'contextual integrity'. Data use which can be completely acceptable in one context, is not so in a different context, even if it involves the same data.
Obviously this is true and the challenge is distinguishing between contexts: which data and which actors, such as senders, receivers, data subjects and purpose of the data use, belong to one context and which to another. Distinctions which can only be made when considering underlying moral but also social values and customs of that context and the interests involved. Hence the outcome might be different over time and between cultures.

Health care as a context would be too undifferentiated for such a nuanced approach. I have mentioned the collective function, but the applications vary widely. Giving care as a professional would be a relevant context. In the present prevailing legal doctrine researchers are excluded from that context. They are seen as alien to this individual function of the use of patient data.
Yet, it is acknowledged that practitioners can and even should use patient data for quality assurance. This is seen as inherent to the individual function. Only in a very extremist view is such a review of data by the practitioner for quality assurance seen as an activity for which the patient should consent. Practitioners often need help by others for quality assurance programs. Criteria have been developed when this quality assurance becomes research, in other words, when the coin flips and the regime of consent would start to apply (Lowrance 2003) They revolve around two aspects. The first is whether the practitioner tries to improve his practice and that of his colleagues according to present standards or whether one tries to find new standards. The second is how many and what others will be involved in this process. Whether they are also involved in the daily practice of care -

even if not with these particular patients earlier, but then at the background of the process such as data managers for the EHCR of the practitioner - or not. Obviously there is a grey area here and the conclusion that the coin has flipped will have an arbitrary ring.
A more fruitful approach is that using data for quality assurance according to present standards and using data to raise those standards are part of the same, along a continuum. Researchers are part of that continuum. It is also by their data that the patient will be diagnosed and treated, that this treatment might improve if new data come in time. Researchers and practitioners will interact at scientific conventions, or, more closely connected to a particular treatment, in small groups which discuss particular patients. They contribute together to multidisciplinary professional standards, are all members of 'learned societies' and share an oath of confidentiality. While as mentioned in section 10.2.3 many new data analysts have been employed in Dutch hospitals to analyse patient data to submit those 'indicator data' of dubious scientific quality, researchers who could provide indicators with sufficient scientific quality are according to present doctrine kept at arm's length from those patient data.
Here it is submitted that researchers belong much more to that context of care than is assumed by the prevailing legal doctrine.

Again, this does not mean that everything goes. The same applies to other contexts. In the context of a circle of friends we can share and are often even expected to share but limited to what we perceive as belonging to what this friendship is about and that will depend on the situation, the purpose and the friends within that circle. But friends still remain different from strangers, which helps to decide which is completely wrong to share, which is automatically shared and which is sometimes subject to an intricate moral balance, such as when a friend is in trouble and too stubborn to accept help. The contention here is that researchers with data are not by definition strangers to the bedside.

10.6.5 Group privacy as an invalid counter argument

So called 'group privacy' refers to the situation that general knowledge about a group becomes available and could lead to discrimination of that group or feelings or anxiety or being wronged amongst its members, even if discrimination by third parties would not follow. Group privacy is often used as an argument against research with data without consent (Callens 1995; Custers 2004; CIOMS 2009; Mascalzoni 2008; Ploem 2010 (the latter two referring to data becoming available through the use of tissue for research).
A bit silly but very true is that withholding individual consent cannot help against data becoming available about a group if others of that group would consent, though the data might be less nuanced. It is even more true that group privacy could also be used against the use of fully anonymous data for research. The use of fully anonymous data is at present not subject to any consent at all.

The argument against using group privacy as a barrier to research with data goes deeper than the previous remarks. Yet, this argument holds only in a democratic society with free press and where the rule of law reigns. Amongst other things people can decide for themselves to which group they belong and often are part of various groups, partially overlapping and sometimes not at all connected to each other. Second, my argumentation is limited to *health* data about 'groups'. Admittedly it might sometimes be more difficult to apply this argumentation to other types of data. I will start with an example to make my case.

The most well known example of group privacy and subsequent discrimination is that of smokers. It shows two aspects of group privacy. First of all, data relating to a group have two edges. Opposite those who feel wronged by those data are those whose lives are saved, because they or others can take preventive measures or arrive at a faster diagnosis. I can see no exception to this 'reception' of health data within the group itself or within health care.

The second aspect is about 'discrimination' of the group outside health care. Obviously smokers are

discriminated, certainly in the non pejorative sense, and some smokers might see this discrimination as in the pejorative sense as well.
However, there will always have been first a translation from general scientific data to measures taken by institutions. There will be a second translation as well, the application of these measures to individuals, rightly or supposedly belonging to that group.
If those discriminatory measures are taken by public authorities, they are subject to public debate and parliamentary control. All arguments have been exchanged in that context and smokers have had their say as well. The measures will delineate how they should be applied to individuals and also that is subject to control.
But what about discrimination by private entities, such as private insurers? That will depend on which private entity and what application. I will come back to that.
First a much more contested example of 'group privacy'. There is a longstanding research cohort of gay men, organised by the Amsterdam public health authority. It investigates the prevalence and incidence of sexually transmitted diseases among gay men and risk factors associated with this transmission.[96] The analyses are performed on blood samples taken from the participants. Of course, participants also answer questionnaires about life style etc. The cohort existed already in the eighties. When the human immunodeficiency virus (HIV) had been isolated, the samples of this cohort were screened on the prevalence of HIV. This was done anonymously and without the explicit consent of the cohort. It proved that many promiscuous gay men in Amsterdam were already carriers of the virus and hence that, even though at that moment there were very few known cases of aids in the Netherlands (Danner 1983), aids would not be a phenomenon restricted to the United States. Publication of this result in a press release[97] led to a small uproar. This shouldn't have been done and if so, the results should first have been discussed with the spokesmen of the gay community More about this episode and how the Netherlands dealt with the HIV/aids crisis in Mooij 2004.
We can see here the two aspects of data relating to groups and the two edges of the first aspect. Opposite those who felt wronged or even in this case existentially worried by those data, there are others, including myself, whose life is probably saved because these data have become available in a timely manner. It is my strong opinion that in this balance the latter weighs more[98].

As far as the second aspect is concerned, group discrimination by private insurers did happen after these findings, as this had already happened to many others based on much cruder data, such as to those who submitted in their application for a life insurance that both parents had died early from a heart disease. (Dis)advantages at our start of life and our later acquired imperfections are regretfully unevenly spread amongst us. It should be constantly debated which distinctions and exclusions may count as justified by private entities. Risk calculation and policy differentiation is inherent to the business of private insurers. They discriminate to keep in business and offer adequate polices. In a civil society private insurers are subject to public scrutiny, to private law and public law. A balance was found which, as with every balance, is always debatable. At this moment, a HIV infection is considered more as a chronic disease. Private insurers do not screen anymore for risk factors to HIV in their questionnaires when someone applies for a life insurance. There is still group discrimination such as gay men being excluded from blood donation. This example shows again that there is public scrutiny and applicable public law mediating between data and group discrimination. Such exclusion has been challenged on several occasions before the Dutch Equal Treatment Commission[99] and upheld based amongst other things on health data about HIV seroconversion of allegedly safe blood donors[100].

It might be argued that I mentioned practices of members of groups and not inherent properties. One can give up smoking. One cannot give up being gay but one can stop with having unsafe sex. But one cannot give up being black and having as a group a higher propensity for sickle cell anaemia. Carrying one of the genes involved had a certain evolutionary advantage in malaria ridden regions. Two is a serious health risk. Yet, the same applies. In this case one cannot oneself take preventive lifestyle measures yet the knowledge will help individual health care by speeding up the diagnostic process. Doctors always discriminate in the non pejorative sense. Without the likelihood

of what disease belongs to this age, gender, race, social class and lifestyle, the diagnostic process would be completely in the dark. Obviously there should also be a balance here, so that the doctor is not blinded the other way. It's about the patient's story first and then connecting it to those broader categories. But those need to be there and they all refer to groups. In this instance this general knowledge might even lead to screening of risk groups, as the results are 'actionable' in the sense of the Wilson and Young (Wilson 1968) criteria. Preconception genetic counselling can be offered. Patients and volunteers belonging to the risk group will form platforms to discuss and promote treatment, education, additional support, etc. [101] There are countless other examples.
The follow up translation of those risk factors outside health care has been discussed already. In this case it is not foreseeable that this specific group propensity would lead to discrimination of the group as such. It might lead to positive measures by public authorities, such as making funds available for screening programs.[102] Private insurers might discriminate when *having* sickle cell disease, just as this happens with other diseases. That can be challenged, either individually or collectively, on the grounds that this disease is treatable and not a relevant criterion. Such challenge is based on the research with patient data.

I fail to see how group privacy can ever be an argument to not seek patterns relating to health risks of certain groups. Instead of promoting the protection of groups, the argument of 'group privacy' (as a barrier to health research), actually is detrimental to the members of those groups. It prevents relevant health data becoming available for members of the group and health care providers.

Yet, the group privacy argument starts from a laudable perspective. It aims to protect members of groups against societal risks to which they cannot protect themselves. It shows distrust in how we can deal as a society with the follow-up translation of health data pertaining to certain groups outside health care. If the situation is that bad, we should combat such social exclusion, not health research.

There is an additional wider argument against group privacy as a barrier to research. That would require a longer discussion, but I will be relatively short on it. Basically group privacy as a barrier to research denies members of the group to have access to data which might be relevant to them. It also denies them and others the opportunity to discuss them. My personal involvement with health data for research, stems from mentioned period of the Amsterdam cohort. If the captains of the gay movement would have had their way, they would have known those data first, kept them secret from the wider public in order to discuss them in a committee with health lawyers and civil servants. The outcome after – at best - several months would be what I was allowed to know. Hence, the group privacy argument shows strong paternalistic[103] tendencies.[104] Yet members of groups to which those data relate, are very well capable of coping with them and incorporating them in their life decisions or interaction with health care providers, as I have tried to show here with my personal narrative and which is corroborated by sociological literature (Rose 2007). One might even consider it anti democratic. Public discussion cannot take place as the data which should be discussed then have become a sort of a taboo because of this group privacy. The formation of new alliances between group members is hampered as the data they would use to form those alliances are withheld from them.[105]

10.7 Health research as a public interest

As seen in chapter 9 the exemptions of Directive 95/46/EC which allow health research with data have been implemented in a vary varied way for health research throughout Europe. The important point here is that except perhaps in the Scandinavian countries (Hakulinen 2011) such research is seen as the exception on the use of patient data instead of an integral part of health care as we know it. This Report and especially this chapter can be seen as a plea for a paradigm shift. Processing patient data for health research is of equal importance to processing for the individual function. It forms part of a continuum.

In the longer run, the update of the Directive should express this paradigm shift. Yet the approach forwarded here can be defended already under art. 8 section 4 of the Directive, which speaks of a 'substantial public interest'. Obviously the interests of patients and the public at large to achieve better treatment and that avoidable risks resulting in diseases will be averted, is substantial to them.

It is also a public interest. In public health care ethics there is substantial discussion about the term 'public' (Dawson 2007) in an endeavour to establish why certain interests are *public* health interests. The conclusion of some of the authors in the debate seems to be that public interests cannot be private interests but relate to "risks we all share" (Gostin 2007) .

I disagree and rely on how (European) public law has settled this discourse already. A public interest is not a yardstick but the outcome of a balanced political decision about what private or societal interests we want or even need to defend. Public interests refer to those specific individual or societal interests which are found to be sufficiently vulnerable and valuable to be protected by governmental measures (WRR 2000, p. 21-22).
Those measures can be many, from coercive, to allowing which would otherwise be prohibited, to supportive such as by making funds available. It may go without saying that interests of future patients are sufficiently vulnerable (amongst many other things they cannot ask for consent from the present patients) and valuable. Researchers mediate, as it were, between present patients and future patients and can in terms of the present Directive use the 'substantial public interest' exception for allowing that which would otherwise be prohibited.

Which brings me to a remark about alleged tension between patient rights and the 'freedom of research'. Though not part of the parlance of the Directive, it is often used in this context (Ploem 2004; Ethikrat 2010). In such general terms patient rights should of course always win. Here we have a specific kind or research, with specific means and specific outcomes. This research is a means to an end. Seen as input and output (with safe throughput) it is not about patients versus researchers but about the data of former patients to get new data to the aid of future patients. As argued, it is then much more difficult to see why the rights of those former patients should win.
In my opinion updated European privacy legislation should express that better, seeing the use of data for health research less as the exception but more as the basis of better health care for all in a solidarity based health care system.

That is also how patient organisations see it, who have recently forwarded 'a right to innovation' which should under circumstances weigh heavier than autonomy about the 'further use' of data or tissue (van der Valk 2010).

10.8 Some practical implications

The basis should be that patient data can be used for health research and the safeguard should not be individual consent but *when* and *how* they are used.
An opt-out option should be offered to all patients. This opt-out should also cover the use of 'anonymous' data for research, as the distinction between indirectly identifiable data and anonymous data has become arbitrary. Both types of data can be used for the same purposes and both will not lead to re-identification of the patient concerned. I will discuss the practical implications in the next part.

In principle only indirectly identifiable data should enter the research domain. Before entering in the research domain (meaning the database(s) to which only researchers have access and where data will be analysed to find the often mentioned patterns) researchers should be able to have access to directly identifiable data in patients records or the EHCR if that is necessary and proportional for this specific research:

- To find patients eligible to enter into a trial (as also argued by AMS 2011) or in a cohort

study with volunteers;
- As a pilot to check whether the patient data are sufficiently suitable for the envisaged research;
- Either in advance for all data or at random at a later stage to check whether data (to be) submitted by the health care provider concerned into the research domain do not create bias (such as entering only those patients in the research domain where the treatment has been successful).

Consent will obviously always be necessary when patients or healthy volunteers are asked to answer questions. One does not get those answers otherwise. Consent should also be necessary when data are accessed outside the present context of research with patient data, being - in short - unbiased research for the public good.

Which will open a host of new questions, coming down to a system of compliance with the above. When does research fall outside mentioned context or when is it actually *necessary* for the research to have access to the directly identifiable data? Some of those questions will be discussed in the following concluding chapter.

11. ANSWERS TO SOME REMAINING QUESTIONS

11.1 Introduction

This chapter takes up some of the remaining questions. Not everything can be discussed. The previous chapter was essentially a plea for a paradigm shift:

- Alongside the individual function the collective function of patient data is just as important for the sustainability of solidarity based health care systems which offer good quality of care;
- Health research is part of that collective function and looks at patients as 'end-products' of health care or whether societal arrangements sufficiently protect us from health threats;
- Using patient data for the collective function is not alien to the privacy context of patient care and present patients may be expected to contribute to the better care of future patients, if certain conditions are met.

The first two aspects are more factual. They explain the present dataflow in health care systems much better than the present 'this is the exception paradigm' does. The third aspect is much more contentious. It also explains certain things better such as the values and professionalism of many researchers which I interviewed, that of many patient organisations, and the dataflow for research in many Scandinavian countries (Hakulinen 2011). It is at odds with much of medical law.[106] That is inherent to a different paradigm.

Even if one would be slightly inclined to support the new paradigm, one can raise many 'yes, but' or 'what if' objections. Such as: ' who will ascertain that this will be good health research?'. It is very easy to find more. I will not discuss them all, but propose a general approach to such questions in paragraph 3.

The 'yes but' reaction generally follows from the following. The 'this use is the exception' paradigm seems to have one major advantage. As it is the exception, the conditions to grant the exception must be established in each and every case. That guarantees that the principles of necessity, proportionality and subsidiary will be applied to such processing by some kind of regulatory body. In theory, this is a strong argument. Yet, these principles are just as applicable in the individual function as well. We have left the application of those principles there much more to the discretion of the health care professionals, without abandoning ex post-regulatory oversight.

For health research with data a similar approach is put forward in paragraph 3. For other applications of the collective function, by governmental bodies or those with a special authorisation to perform functions in the health care system, such as health insurers, the general principles of public law apply. Any authorisation to act must have a basis in a statute or by-law. That also holds for processing data. There is, as argued, a lot of such processing yet most of it does not involve personal data. Such bodies cannot be seen as part of the context of care. They have different interests. Any use of personal (hence) patient data should indeed be the exception.

11.2 Data are anonymous if the conditions mentioned are met

Directive 95/46/EC uses a contextual approach. Recital 26 mentions 'the means likely reasonably to be used to identify'. Little has been left of that contextual approach in the interpretation of the various DPA's. The art. 29 Working Party (Working Party 2007) admitted that there is a grey area between still identifiable data and anonymous data and argued that it would always be better to err on the side of considering data still identifiable. In that case they wouldn't escape the scrutiny of the DPA's and the application of the Directive. The working party stated that national laws of the member states allow for sufficient flexibility to use those data for legitimate purposes, such as research. Researchers in many member states on the other hand encounter more rigidity than flexibility. These problems for health research will increase if more and more datasets are considered identifiable.

The proposed solution brings back the contextual approach, yet without abandoning regulatory

oversight completely. They can be considered anonymous data only if those conditions are met. How it will be established whether these conditions are met, is, as argued in Appendix 1, left to national law. Hence there is still regulatory oversight to ensure that the data-handling by research entities will be secure.
But if so, it cannot be reasonably argued that those data which are transmitted from the data sources and are being used (to avoid the term processing, which relates in the Directive to personal data) in the research domain are still personal data. They are anonymous data from the moment of their release by the data sources.

11.3 Additional questions

11.3.1 In general

As mentioned, the arguments given in the text can raise a host of objections. I can not discuss them all. A distinction is helpful into two categories of possible objections and a general discussion of the possible objections within each of these. These are:

- How can it be ascertained that researchers will handle the data responsibly for the research mentioned for the public good and not for other purposes;
- I have pleaded for a paradigm shift in part 3. Inherent in the function of health care is to use patient data for research. Yet, it should be possible to opt out. I propose a solution to the present problems for health research which is obviously influenced by this alternative paradigm, yet there is no-opt out in most member states for the use of anonymous data for research.

The first type of objection is more practical. It will be discussed in the remainder of this section. The second objection is more principled. That will be discussed in the concluding paragraph.

In most member states many safeguards exist before non anonymous data can be released for research, especially if this would be without consent. They range from statutory provisions that a certain type of data should be transmitted to certain registries (such as in countries where cancer registries are based on statutory provisions) to the approval of an ethical review board in each of the centres (data sources) which would release such data to the research domain and the approval of the DPA. We surely would have eliminated all possible 'risks' in the latter case, though that comes with huge costs. Also when one would consider the latter system of oversight disproportionate, every system seems to be circumvented by considering such data anonymous.

If this system was meant to assure (sufficiently) that data would not be illegitimately re-identified, it is obviously not necessary anymore. That has been secured in a more structural way. Instead of judging each project, the research entity (see Appendix 1) has been approved for all its projects on this issue of data security.

If the system of oversight is meant to judge which research may be performed based on criteria such as whether this is 'good research' or whether the research really serves a public interest, or even a 'substantial' public interest etc., this Report faces a different challenge. Discussion of such topics can be very lengthy and would lead us too far away from the scope of this Report.
I would like to make the following perhaps contentious remarks which should be sufficient for the time being.
The first is a general remark about regulatory systems of oversight. Why do we want it and do we really need it?. My contention is that many want it because they trust themselves better than everybody else (in this case as members of a –bureaucratic- system of oversight better than professionals) but that we only need this when the risks (both in terms of the chance that it will happen and the severity of what will happen then) are really serious (see also Sunstein 2005).

Connected to this the following. Good health research with data is always in the public interest.

The basis of such 'good health research' is research which is methodologically sound and where the results will be published, even if they are inconclusive. Such research is the remit of the research entities discussed in Appendix 1. They are externally funded for nearly all specific projects. The judgement whether this project will have added value for health research is primarily made by those external funders. In nearly all cases those are not-for-profit organisations. I don't see the need to re-evaluate the project on these aspects somewhere else, and instead may point at what has been said about the 'professionalism' of health researchers in chapter 9.
The same holds true for the 'what if' when the third party funder, or possibly co-funder, would be a pharmaceutical company. One of the basic elements mentioned is that the results should be published. That should be the case even if the financer(s) of the project does (or do) not like those results. The contract which establishes the funding should be clear about that.

Much more could be said, such as about 'conflict of interests' policies, an obligation to datasharing, and the (other) good research governance principles I have mentioned elsewhere (van Veen 2008). The interesting point here is that they relate to the functioning of the research entity as such. They have to implemented there, and would only need fine tuning at times for each project. They do not need to be assessed for each project separately by some or even several review bodies.
Yet, it will remain the judgement of the data sources whether they will submit data for the project, insofar as the project is not based on a statutory obligation to submit data. Securing the datachain as described in Appendix 1 entails amongst other things a data transfer agreement (DTA) with each of the sources.[107]

11.3.2 Direct access to patient records

The new paradigm allows for direct access to patient records , yet this should still be the exception. Hence the safeguards of the previous section are insufficient for such special use. Here a case-by-case review would be appropriate. How that will be organised is beyond the scope of this Report.

11.4 The way forward

As argued in chapter 10 there should be the possibility to opt out, whether data are still personal data or anonymous data, unless the data is to be submitted to a registry or the research domain based on a statutory provision.
That is the corollary of the paradigm shift put forward here. The proposal is obviously not meant for 'cherry picking', leaving the present system for the use of (indirectly) identifiable data as it is and adding the opt-out for anonymous data. Whether the paradigm shift will be embedded in the new Data Protection Directive and if so how, is a discussion which will continue for some while and cannot be settled here. It is more than a legal or ethical discussion. It is also about regulatory traditions and the interests of those who have a stake in the various present systems of oversight in the member states.

The opt-out is based on the following more principled arguments:

- It can help ensure that the proposed system will not endanger access to health care (chapter 10);
- The distinction between indirectly identifiable data and anonymous data has to a very large extent become arbitrary (chapter 9).

There is also a more practical argument. The Netherlands have research exemptions to the consent principle. These can only be used if the patient did not opt out. I have been involved in various information brochures for patients or possible participants in screening programs. It is virtually impossible to draft a concise text – on the educational level for which these brochures are meant – which explains that:

- Anonymous data will be used for research without your consent;
- If non anonymous data will be used, you will be asked for consent;

- However, this is not always feasible and then non anonymous data can be used, but your privacy will nevertheless be sufficiently protected (researchers will only see indirectly identifiable data and nobody else will);
- But if you don't want this, you can object to this.

And sometimes it is unavoidable that patient data are used for registries which, by the present paradigm, are seen as outside the privacy context of care. This is especially the case during pregnancy and shortly after birth. The several screening programs and scientific programs aiming at raising quality of ante and postnatal care, amount to a sort of 'multiple choice exam' about data use for the women concerned. We diligently try to describe that in various brochures, which lead some women in a focus group reading them exclaim: 'why all these legalities, where is the bliss of pregnancy and child birth' (the sweet Dutch expression is literally translated: 'my pink cloud').[108]
This is again an example of seeing patients, or in this case not patients but pregnant women and young mothers, not as real persons but as abstractions which must fit our normative ideas of how they should interact with health care (see section 10.3).

It is much more understandable for them and in the light of the above arguments also more fair, to simply state that health research with data is very important. You have profited from earlier research with data, other patients will profit from new research with data. Also data about your treatment can be used for such research. We will protect your privacy during this research. Your data will not be disclosed to parties which do not strictly need access in the course of this research and who are subject to an oath of confidentiality. Nobody else will have excess. But if you don't want this research to happen, you can object to this. Of course, that will not in any way influence your present treatment.

The nature of any opt-out or objection system is that the patient should take the initiative. Yet, there must be sufficient transparency and a low threshold to opt out. I cannot go into the details here. The challenge is to develop good practices in this respect without them amounting to a de facto opt in and hence, consent system.

The short, rather loose text mentioned above might seem to show a bias for research. In a Patient Information Form (PIF) in a clinical trial such bias would be unacceptable. Yet, this is not a PIF. The stakes are different here and so are the ethics behind allowing use of data instead of participating in a trial. Such bias here is completely acceptable and addresses patients as responsible citizens.

A last general remark about this system. It will take some time to implement it. It must be implemented at the data sources, the health care providers. Again this requires procedural and technical measures. The possible opt-out must be noted in the EHCR's once and follow the whole chain of events with the patient data. Few EHCR's allow this at the moment.
From a regulatory point of view it would be helpful if the system is not only laid down as a condition for research with data (not if the patient opted out) but also as a requirement for health care providers to implement the opt-out.

APPENDIX 1

CONDITIONS FOR CONSIDERING DATA ANONYMOUS IN THE RESEARCH DOMAIN

1. The starting point: data which are not directly identifiable

This Appendix describes (until section 5) the conditions under which data which are not directly identifiable can be considered anonymous data in the research domain.
Not directly identifiable means that names and addresses or the civic registration number are removed and have been replaced by a number to uniquely distinguish those in the database from each other. That number is applied to the dataset somewhere earlier in data-chain, before the data entered the research domain. It might be a random number attached to the research data by the source or a pseudonym based on an algorithm. The latter means that the procedure can repeated by a data source and that the same person will again get the same pseudonym. The pseudonym might be irreversible or reversible. In the first case the holder of the key to the algorithm (at the data source) cannot go back from the pseudonym to the direct identifiers. In the latter case the key holder can.
The first challenge of regulatory authorities is that this unique 'number' or 'mark' (to avoid the term 'identifier') to distinguish the data relating to different persons from each other is not sufficiently secure to prevent retrieving the identifiers behind that number or mark.

The second challenge is that the research data attached to that unique 'number' or 'mark' of the persons involved, can lead to re-identification by matching them with other data which are directly identifiable and publicly available. That challenge has become more acute since the proliferation of personal data on the internet.
These research data often need to be very detailed about those persons. Address is one of the direct identifiers but can be a relevant research data as well. The same applies to date of birth. All together that can make these research data indirectly identifiable concerning the (all or some) persons concerned. The usual answer is further aggregation of the research data. Instead of full zip Code only three digits, instead of date of birth 5 year classes, etc. As has been shown in part 2, this quest for complete anonymity regarding the aggregation level is a lost battle. It would lead research into a data desert, if data need to be relevant for the more detailed research which epidemiological research often is.

The proposed solution is not the quest for anonymity before data reach the research but safeguards against re-identification once they are there.

2. The two basic rules: a data use policy and technical measures keeping track of what has been done with the data in the research domain, by whom and when, alongside external security.

This is not the place to repeat norms which have been described elsewhere on data security and confidentiality, such as the relevant ISO norms[109], the Dutch norms in the NEN 7510 series[110], or those described in the 'Guidelines on confidentiality for population based cancer registries in Europe'[111]. This appendix would become very long in that case. It should be sufficient to mention the basic principles. These principles hold for each 'research entity'. Those research entities have to implement the principles, using the documents mentioned. With the term 'research entity' I refer to an entity with its own remit and responsible head with sufficient (delegated) authority, budget and discretion to implement decisions at the level of the entity, even if that entity forms part of a larger whole, such as the department of epidemiology or biostatistics at a University Hospital.

At the level of the 'research entity' there should be:

- A written policy about the use of data submitted for research, signed by the 'director' or head of the entity;
- This policy should stress the confidentiality of the submitted data and the fact that the

data should (only) be used according to the specific purpose for which they were submitted, however allowing for approval of different possibly fruitful public health outcomes of the research if the serendipity inherent to any research[112] and specifically research with data[113] would lead to that;
- The authorisation of which researchers will have access to which data;
- The external security, referring to measures ensuring that non authorised persons (either within or outside the research unity) will not have access, and this over the whole course of the data chain;
- The internal security, referring to measures that data will only be used by those authorised according to their authorisation.

The last point is the most important for the present discussion. If this is implemented, it means that the persons involved will not be re-identified, or if this has ever been the case, or could have been the case, this will come to light soon enough and appropriate measures will be taken.
At the same time this poses at this moment the most practical problems for implementation. In practice it means logging and the possibility of an audit trail of the data showing who has done what with the data and when, including making a copy or merging with other data outside the research database. From my interviews with researchers I have understood that many of the databases on which researchers work at the moment, do not have this technical feature. It is a separate database with sufficient external security but with software designed to make all these statistical analyses, aiming at validity of the results. The software is not designed for this kind of audit trail.

3. Practical interim proposal

It will take some time to achieve the kind of databases which are also secure from illegitimate use by researchers authorised to use them. The 'terms of reference' (TOR) for this software have to be rewritten first. I would like to add: and without making adaptations to present software excessively expensive.

For the time being it is suggested that research entities which have proven to handle many data without ever having an incident happen, can submit a statement that data are safe at that centre, meaning both:
- safe from attempts to gain access by anyone not explicitly granted access to the data;
- will only be processed according to the research protocol for this study and (hence) are safe from re-identification by those who have been granted access

The conditions for such a statement would be the following:
- The centre has a system for quality assurance and the safe handling of the data is part of this QA system (though as argued that system cannot as yet be as fully developed as described earlier);
- This involves, amongst other things, secure transmission of data, job descriptions, SOP's for data handling and logging of procedures as far as possible with the present software tools;
- Each dataflow can be based on a protocol which explains why those data are necessary for a specific research project;
- The number of staff members who have access to the data is limited to those who really need to work with those data. Often various projects are taking place at the same time at the centre. Each project should have its own secured database (though that can be on one server) and access procedures to that database;
- There is a separation of functions between staff members who perform administrative tasks and in that function have contacts with the sources which submit data, and those who apply pseudonymisation procedures and those who analyse the data to arrive at results of the research. It must be acknowledged that often the scale of the research entity or of the project is too small to allow those functions to be assigned to different persons. In that case it must be clear in what capacity that person has acted and that rights assigned to one function cannot be used in another function. E.g.: the research assistant who is contracted for this project (thanks

to the grant) first negotiates with the sources about the data and only in a later stage uses data for the analyses. Access to the coding mechanism which was agreed with the data sources will not be possible in that later stage;
- All persons who handle data have signed a confidentiality agreement (often this is already part of their contract, but it needs to be stressed explicitly);
- Most of them are epidemiologists registered at their professional organisation. That assures some system of professional oversight. Dutch epidemiologists are subject to the Dutch Code of Conduct on health research, which explicitly forbids re-identification;
- There is a pledge that serious breaches of the policy, such as data loss to third parties or illegitimate re-identification by staff members will be reported to the relevant regulatory authorities even if national law does not require this (yet);
- The statement is signed by the director/head of the research entity and that director should be a professor. This assures that someone is really responsible and accountable.

When those conditions are met, indirectly identifiable data – according to that very broad conception – can be considered anonymous data at the centre concerned.

4. Remaining questions

Starting with the last proposal, the following questions remain:
- Who are those research entities which have an untarnished track-record for handling data?

This can only be answered on the national level. In the Netherlands all departments of epidemiology (often they have different names) of University Hospitals would fall into this category and, without being exhaustive, independent research institutions as NIVEL, TNO, IKNL for the aggregate of the cancer registry.
Admittedly, this places traditional entities in a better position than those entering the field. To them the proposal described in section 2 would apply. Both for section 2 and section 3 the following questions remain:
- Must an authority, such as the national DPA, first ascertain that the policy or statement is there and grant an official recognition?
- How will oversight be effectuated to see that the policy is effective in practice?

Again these questions can only be answered on the national level, in correspondence with the national regulatory traditions and implementation of Directive 95/46/EC.
Regarding the first question, in a tradition where every use of research exemptions has to be approved by the DPA, such as in France[114], it seems logical that the policy has to be submitted to the CNIL (the French DPA). In a less bureaucratic tradition which also relies more on self-regulation, such as the Netherlands, it seems logical that it would be sufficient that this is embedded in the self-regulation, or Code of Conduct, which is approved by the DPA.[115]
The second question relates to the following. The policy and ensuing measures, both procedural and technical, shouldn't just be there on paper but must be effective and function according to their intentions as well. It must be checked whether that is the case. There are several options here, ranging from less to more costly. These options are not mutually exclusive:
- There is regular oversight by the staff members using checklists etc (described in the policy) how the practical compliance is effectuated;
- A staff member is exclusively dedicated for the job as 'data security officer' and has certain competences for oversight in this respect;
- Regular audits by certified third parties;
- Regular audits by regulatory bodies.

It might not be surprising that I am not in favour of the latter two options as long as there is no need to. Both can be very costly. Yet it has to be acknowledged that the scale of certain projects (in time, amount of data, complexity of the data chain) or the scale of the research entity can be such that external audits from time to time are helpful to ascertain the proper functioning of these. If the research has its data handling policies and implementation well organised, such audits could also be less costly than is often assumed. It must be stressed that the audit should not be about

whether the persons involved in the data are really anonymous but about what has been described here: external safety and internal safety against illegitimate re-identifying.
With such external audits I do not see a function for regular audits of regulatory authorities. Submission of the audit certificates should be sufficient if national law would require close supervision of the research entities.

5. Application to processing of identifiable data correspondingly

All research entities have several projects running. Many of these involve processing of personal data such as with panels of volunteers. All research entities I know of in the Netherlands make a distinction between those involved in the administrative tasks of contacts with respondents, the reminders etc. and researchers who work with the answers and other data, if so approved by the respondents, from data sources such as cancer registries etc. As such, these researchers do not have access to the direct identifiers of the respondents. Promises are made to the respondents in this respect: their data will be processed 'anonymously' or more to the point, given the present state of the law, your identifiable data will not be disclosed to the researchers.
The procedures proposed here, should apply to all the processing or handling of data at the research entity. In the present case not to ascertain that the research entity does not process personal data for this project but to ascertain that the promises mentioned are an accountable reality. One the one hand this will make the policy and technical procedures slightly more complex. They should differentiate. On the other hand, it is not very practical to leave certain data processing out of the procedures. External safety applies to all processing. Internal safety as well. The differentiation is which staff members can do what. Certain staff members will have access to certain personal data in their function as administrative liaison with the respondents.

[67] Privireal states the conception differently, namely being "personal data about them (data subjects) that can negatively affect their physical, psychological and moral integrity". Stated as such it lacks any distinction with the narrow conception. See also Chapter 10.

[68] See the overview in Beyleveld 2004.

[69] See www.eurocourse.org

[70] Mainly due to lacking exclusive jurisdiction in a particular field of labour, at least not in the literal sense of jurisdiction such as lawyers (when registered at the bar) or physicians have.

[71] For brevity's sake I will use the male form here where also the female form should be read.

[72] For a fuller description see Ministry of Health, Welfare and Sport: http://english.minvws.nl/en/themes/health-insurance-system/default.asp (last accessed 10-04-2011).

[73] Art. 81HciA.

[74] Previously this type of care was reimbursed through another collective system not discussed here.

[75] CBB 02-08-2010 LJN: BN 3056.

[76] Art. 60 and following Wet marktordening gezondheidszorg (Act on the healthcare markets).

[77] CBB 11-2-2010, *TvgR* 201/12.

[78] In the sense of ISO/TS 25237: 2008 Health Informatics-Pseudonymization.

[79] See: http://www.dbcinformatiesysteem.nl/Home.

[80] Wet op het Centraal bureau voor de statistiek (Act on Statistics Netherlands) 2003, art. 34 and 35. See also the Ministerial Decree on causes of death, (2007) to be submitted to Statistics Netherlands.

[81] This is based on art. 5 of the Quality of care of health care institutions Act and art. 36 of the Act on health care markets, in combination with practical agreements by all stakeholders including patient organisations on which 'indicators' should matter in a given year. Needless to say that the number of indicators is expanding each year.

[82] Prof. R. Brand on a conference on "Indicatoren met één click" (indicators by one push on the button) by Medical PHIT on September 28 2011.

[83] I am quite aware that this list partially overlaps. We might acquire an infectious pathogen from the environment and then transmit it to someone else, as it has happened with the bird flu. The reaction to a certain pathogenic factor can be influenced by the person's (genetic) propensity to deal with it. An infection with the hepatitis B virus might go unnoticed to some while others get very ill from it, and even developing liver cancer. Some of these patients might react positively to antiviral treatment, while a few others might not. This is, by the way, a reason why we should know more about patient data, not less.

[84] Meaning, diseases transferred by living organisms *and* which can be transferred from one person to another.

[85] The basis is Decision No. 2119/98/EC of the the European Parliament and of the Council of 24 September 1998 setting up a network for the epidemiological surveillance and control of communicable diseases in the Community, *OJ* L 268, 3 October 1998, p. 1–7. A very large number of Commission Decisions have followed from that. For a full list see www.ecdc.europa.eu /en/aboutus/key_documents/Pages/key_documents.aspx.

[86] They might already be carriers and have infected others, or are still infectious, but may not have noticed this because of their better general health and response of the immune system. Tuberculosis is a case in point.

[87] See for the Dutch situation amongst other things Handboek patiëntenparticipatie in wetenschappelijk onderzoek, ZonMW , Den Haag, 2006 or the many activities announced on the site of the VSOP www.vsop.nl or, outside the Netherlands, its sister organisation Genetic Alliance UK, http://www.geneticalliance.org.uk/

[88] And not only in the Dutch discourse. See as early as Schloendorff v. Society of New York Hospital 211 N.Y. 125, 105, N.E. 92 (1914) with Justice Gardozo famous quote "Every human being of adult eyars and sound mind has the right to determine what shall be done with his own body"

[89] see http://www.cdc.gov/hpv/.

[90] It is difficult to find a clear basis for this in textbooks on medical ethics, like the once seminal textbook on medical ethics, Beauchamp (2009) But the absence of any 'quid pro quo' is in my opinion essential to giving health care. A more extreme example: patients will receive organ transplants when they had not opted-in, or even if they opted-out, depending on the national legislation, of the national organ donation scheme. For a defense of this see Buijsen 2008. Yet, I doubt whether that complicated approach helps us much further in this instance. I found Sen's, in a way 'postmodern' account of justice, (Sen 2009). more helpful. This – helping 'free riders' when they are in distress– would be the *nyana* kind of justice as described by Sen. Yet, it is quite another

thing that we should assist free riders to get that free ride.

[91] The 'tragedy of the anti commons' was as far as I know first mentioned in the context of health care in a paper on patenting human genes (Heller 1998)

[92] Such as the UN Covenant on Civil and Political Rights, article 7.

[93] Such as The Universal Declaration on Bioethics and Human Rights, adopted by UNESCO in October 2005, but the earliest and most prominent, the World Medical Association's Declaration of Helsinki, 1964, last amended 2008.

[94] For the history and origin of medical research and its regulation, see Bergkamp (1988), Brody 1998. Faunce 2005, page 167 and following; Plomer 2005.

[95] And hence in my opinion the doctor might use some soft paternalistic pressure to choose otherwise, van Veen 2003

[96] See: http://www.gezond.amsterdam.nl/Wetenschappelijk-onderzoek/Infectieziekten/Informatie-voor-deelnemers/Cohort-onder-homoseksuele-mannen.

[97] As far as I know, these results have never been published in a scientific journal. A follow-up study with nearly the same cohort and proper consent was published in Godfried 1987.

[98] I have discussed an earlier version of this section with two friends who did not run through the bullets during that time (this expression comes in this context from Mordden 1997) but have survived and a much younger friend who is participating in the present cohort. Though they were a bit worried about my 'outing', they agreed with my conclusions. One of the older friends did not agree with my presentation of exclusion of gaymen from blood donation, discussed in the text infra.

[99] See for this commission: http://www.cgb.nl/english.

[100] CGB 2009-137 and CGB 2007-85.

[101] Such as in this example the sicklecell society in the UK. See www.sicklecell.org

[102] Of course, one cannot offer such a program as a governmentally backed program without solidarity based health care which offers treatment to those affected.

[103] The philosophical literature distinguishes between 'weak' and 'strong' paternalism. Weak paternalism means in short, that a person can be protected against choices which are not autonomous. Strong paternalism means in short, that another person makes those choices for an autonomous person, for example by withholding information enabling a choice. Some weak paternalism is usually admitted in medical ethics; see amongst others Beauchamp 2009 p. 206-216.

[104] Which does not mean to say that researchers don't have certain obligations about how to present outcomes of research, involve patient organisations, etc. I can not discuss this topic here. See Van Veen 2008 for some principles of 'good research governance'.

[105] Practical examples of negative effects of group privacy are rarely mentioned in the literature. An exception is Callens who quotes a hypothetical case which was presented at a CIOMS conference (Callens 1995, p.457) I will quote this example in an attempt to make the case once more. All the elements mentioned in the text come together here. The case concerned 'research ' at a particular school where it was found that a large percentage of the pupils there had started to use illegal drugs. Later in life that might be used against them, whether they had used those drugs or not. Such a scenario seems unlikely to me unless the school was notorious for having 'problem pupils' anyhow. But having said that, the following comments on the relation with the kind of research meant in this Report.
One may first of all question whether this is health research as discussed here. It is applied research relating to a specific case. It is not good research and not health research if it does not indicate the conditions under which those kids started to use the drugs, correlation with mental status, well-being at school etc and hence possible means for prevention (as the alternative to repression). It is not good research as it only relates to one school and simply states facts instead of the relation with the broader picture of drug use under adolescents. That school might not at all be an exception. As such the report might have been submitted to the school but not be published in a serious paper and certainly not with the school's name attached. In the Netherlands all regional public health authorities monitor at regular intervals life style, well-being and health threats of children (where in the Netherlands we have a somewhat different conception of what drugs are illegal, but smoking too much pot for an adolescent is never good of course). The results are given back to each school, anonymously (for the children concerned) and benchmaked against other schools (with relations to socio-economic status, cultural background, etc.). Results for individual schools are never published. That is good scientific public health monitoring and is the kind of research referred to in this Report.
But to continue on this example and to end this 'group privacy' myth once and for all, additionally the

following. It cannot be excluded that a report about one school performing badly, leaks to the press. Recently serious problems at a school for higher education were broadly discussed in the Dutch press. These problems had nothing to do with drugs or health problems there, but with something which can be legitimately used by possible employers against the former students (former drug use cannot). Namely that school had, amongst other things, been much too lenient with its exams and many pupils had diplomas for which they had not yet been ready. Should the inquiry about the educational level of the school not have taken place because of the group privacy of former students? Nobody in his right mind can argue that. Should it have been kept out the press? We have a free society with free press. It helps to correct the behaviour of the board of that school (all the boardmembers were dismissed) and creates a public discussion about this type of school in general. It warns prospective students not to go there. Alumni of the school have formed groups to show that they nevertheless acquired the necessary level of education.
Basically the same applies to the hypothetical case of excessive drug use in one school which leaks to the press. I agree with the promise of secrecy by the way. We want the schools and pupils to cooperate. That can only be achieved by this promise. Hence I would oppose detailed information on specific schools being submitted to the press on request based on our 'Freedom of information of public authorities Act' (in Dutch: Wet openbaarheid bestuur). Yet, it cannot be excluded that an administrative Court will grant the request. The school's name cannot be hidden in that case, the pupils will still appear as statistical numbers. As said, we will oppose and the more likely scenario is that an angry teacher leaks, tired of the excessive drug use of his pupils and the lack of any preventive measures. Something must change at that school and if the school hasn't a good defence for its bad performance, this pressure might help. If I were that school, I would have been open myself, from the parents and the pupils. "We have a serious problem and we are going to solve this". That might very well attract the attention of the press. The school might state that it has called for the police many times to prevent 'dealing' of drugs around the corner or that the Inspectorate for Education had not allowed it to remove certain pupils which are the cause of all the problems. A public discussion will start. The contrary would be anti-democratic and the ugly situation might continue. That weighs much heavier than that very hypothetical situation that it is used against a pupil who comes from that school. Our pupil might even align with the 'drug free pupils of school X' group, which of course can only be formed if the situation is in the open. And even if he or she would be part of the group 'pot has fertilised my creativity' I personally might still hire him or her - ceteris paribus -, even though I hate the very smell of the stuff. Such creativity and thinking beyond the established boundaries seems to be lacking in the debate about group privacy.

[106] Though in its effects this is not very dissimilar to what Ploem had proposed for the Dutch system namely an opt out for the use of indirectly identifiable data for research without the need to ascertain first that consent cannot reasonably be asked etc. (Ploem 2004). We have arrived at this via completely different routes though such as that I consider research with data much less exceptional and that in her conception of coded data, they are always indirectly identifiable. I am more lenient towards the use of directly identifiable data under specific circumstances, namely also based on opt-out. I am in a way more restrictive about the use of fully anonymous data, namely allowing for an opt-out. Yet, following from the arguments forwarded in this Report, this category encompasses much more data, if not nearly all, for research, than when they would still be considered indirectly identifiable.

[107] These can be rather short. My approach to those 'legalities' is to have three tiers, which can refer to each other. The first is the general policies of the research entity, which will encompass the elements described in Appendix 1. The second is the description of the specific project which can refer to the first tier for the general aspects of research by the research entity. The third is then the DTA which can be concise as it can refer to the documents of the first two tiers. The second tier does not need to provide much additional work for the research entity either. Usually there is a protocol already which was written in order for the project to acquire funding. That can be the central element of the tier 2 documentation.

[108] The results of this focus group have not been published.

[109] Such as SAS70, in conjunction with the applicable norms of the ISO 27002 series

[110] NEN 2011. The relevant ISO norms on electronic processing of patient data have been brought together in these documents. However they only relate to processing in the context of patient care, the individual function as described in this Report. They cannot be applied one to one to processing in the research domain.

[111] ENCN 2011. I am referring here to paragraph 5 of version 4 of the draft Guidelines.

[112] What if A. Flemming had not further investigated why the staphylococci were being extinguished in a dish which had accidently been contaminated with a fungus, because he was not 'allowed' to do so as not being the subject of his research?

[113] See Vandenbroucke 2010

[114] See van Veen 2006.

[115] See art. 27 of Directive 95/46/EC

LITERATURE

Agledahl 2011	K.M. Agledahl, R. Førde, Å. Wifstad, 'Choice is not the issue. The misrepresentation of healthcare in bioethical discourse', *J Med Ethics* 2011 37: 212-215, DOI 10.1136/jme.2010.039172
Al-Shahi 2005	R. Al-Shahi, C. Vousden & C. Warlow, 'Scottisch Intracranial Vasculkar Malformation Study Steering Committee. Bias from requiring explicit consent from all participants in observational research: prospective, population based study', *BMJ* 2005, 331 (7522), p. 942
AMS 2006	*Personal data for public good: using health information in medical research*. A Report from the Academy of Medical Sciences. London: The Academy of Medical Sciences 2006
AMS 2011	*A new pathway for the regulation and governance of health research.* A Report from the Academy of Medical Sciences. London: The Academy of Medical Sciences 2011
Appels 2007	C.W.Y. Appels, 'Vertekening van de resultaten door de methode van 'informed consent' bij medischwetenschappelijk onderzoek', *NTvG* 2007, p. 681-682
Beachamp 2009	T.L. Beachamp & J.F. Childress, *Principles of biomedical ethics (6th ed)*, New York/Oxford: Oxford University Press 2009
Bergkamp 1988	L. Bergkamp, *Het proefdier mens, de normering en regulering van medische experimenten met mensen*, Alphen aan den Rijn: Samson Uitgeverij 1988
Beyleveld 2004	D. Beyleveld et al (ed), *Implementation of the Data Protection Directive in Relation to Medical Research in Europe*, Aldershot/Burlington: Ashgate 2004
Beyleveld 2007	D. Beyleveld, 'Data protection and genetics: Medical research and the public good', *King's Law Journal* 2007, 18, p. 275-289
Bilsen 2007	P.M.A. van Bilsen (thesis), *Care for the elderly. An explanation of perceived needs, demands and service use*, Maastricht: 2007
Borking 2010	F.J.F.M. Borking, *Privacyrecht is code: over het gebruik van Privacy Enhancing Technologies*, Deventer: Kluwer 2010
Bovenberg 2006	J.A. Bovenberg, *Property Rights in Blood, Genes and Data: naturally yours?*, Leiden/Boston: Martinus Nijhoff 2006
Brody 1998	B.A. Brody, The ethics of biomedical research: an international perspective, New York/Oxford: OUP 1998
Broeren 2011	J. Broeren, 'Goed geïnformeerde e-patient is fictie', *Medisch Contact* 2011, 66 (22), p. 1418-1419
Brownsword 2008	R. Brownsword, *Rights, Regulation, and the Technological Revolution*, Oxford/New York: OUP 2008
Buijsen 2008	M. Buisen, *The meaning of 'justice' and 'solidarity' in health care*, in Den Exter 2008, p. 51-61
Callens 1995	S. Callens, *Goed geregeld? Het gebruik van medische gegevens voor onderzoek*, Amtwerpen/Apeldoorn: MAKLU 1995
Campbell 2011	A.V. Campbell, 'Autonomy revisited – a respons to H. Haker' *Journal of Internal Medicine* 2011, DOI 10.1111/j.1365-2796.2011.02370.x, p. 380-382
Carinci 2011	F. Carinci et al, 'Revision of the Data Protection Directive: opportunity or threat for public health monitoring?', *Eur J Pub Health* 2011, DOI:10.1093/eurpub/ckr100
CBP 2005	J.A.L. Krabben, Landelijke Zorgregistraties, Den Haag: CBP, 2005, bij het Voorwoord door J. Kohnstamm
CBP 2009a	Onderzoek van het CBP bij het College voor Zorgverzekeringen m.b.t. de AWBZ-brede zorgregistratie, 20 maart 2009, z2008-00292
CBP 2009b	College Bescherming Persoonsgegevens, 'ASP's in de zorg', brief aan de Minister van VWS

CCMO 2009	Centrale Commissie Mensgebonden Onderzoek, mededeling over codering van onderzoeksgegevens, 18 oktober 2009
Chalmers 2011	D. Chalmers, 'Are the research ethics committees working in the best interests of participants in an increasingly globalised research environment?' *Journal of Internal Medicine* 2011, DOI 10.1111/j.1365-2796.2011.02370.x, p. 392-395
CIOMS 2009	Council for International Organizations of Medical Sciences, *International Ethical Guidelines for Epidemiological Studies*, Geneva: CIOMS 2009
Commission 2004	Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions on *Modernising social protection for the development of high-quality, accessible and sustainable health care and long-term care: support for the national strategies using "open method of coordination"*, COM (2004) 304 final
Commission 2007	Communication from the Commission to the European Parliament and the Council on *Promoting Data Protection by Privacy Enhancing Technologies (PETs)*, COM(2007) 228 final.
Commission 2010	Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions on *A comprehensive approach on personal data protection in the European Union*, COM(2010) 609 final.
Curren 2010	L.Curren et al, 'Identifiability, Genomics and UK Data Protection Law', *European Journal of Health Law*, 2010, 17 (4), p. 329-344
Custers 2004	B. Custers, The power of knowledge: ethical, legal and technological aspects of data mining and group profiling in epidemiology, Nijmegen: Wolf Legal Publishers 2004
Danner 1983	S.A. Danner & R.A. Coutinho, 'Het verworven immunodeficiëntiesyndroom (AIDS)', *NED Tijdschr. Geneeskd*., 127(19), 1983
Dawson 2007	A. Dawson & M. Verweij (editors), *Ethics, prevention and public health*, Oxford, New York: Oxford University Press 2007
Den Exter 2008	A.P. den Exter (ed), *International Health Law. Solidarity and Justice in Health Care*, Apeldoorn/Antwerpen: Maklu 2008
Dute 2008	J.C.J. Dute, H.J.J. Leenen & W.R. Kastelein, *Handboek gezondheidsrecht deel II, gezondheidszorg en recht*, Houten: Bohn Stafleu van Loghum 2008
Duthler 1998	A.W. Duthler, *Met recht een TTP!*, Deventer: Kluwer 1998
Eder 2007	M.L. Eder et al, 'Improving informed consent: suggestions from parents of children with leukemia', *Pediatrics* 2007; 119; e849-e859, DOI: 10.1542/peds.2006-2208
Ethikrat 2010	Deutscher Ethikrat: *Humanbiobanken für die Forschung*, Berlin 2010: Deutscher Ethikrat
Faunce 2005	A. Faunce, *Pilgrims in Medicine: Conscience, Legalism and Human Rights*, Kluwer Law International 2005
FEDERA 2011	Gedragscode Goed Gebruik (Gedragscode verantwoord omgaan met lichaamsmateriaal ten behoeve van medisch-wetenschappelijk onderzoek) 2011
Fisher 2011	H.R. Fisher, C. McKevitt & A. Boaz, 'Why do parents enroll their children in research: a narrative synthesis', *JME* 2011, doi: 10.1136/jme.2010.040220
Freidson 2001	E. Freidson, *Professionalism, the third logic*, Cambridge: Polity Press 2001
Godfried 1987	J.P. Godfried,G. van Griensven, R. Tielman et al. 'Risk factors and prevalence of HIV antibodies in homosexual men in the Netherlands', *Am Jour. Epidemiology* 125(6), 1987, p. 1048-1057
Gostin 2007	L.O. Gostin & L. Stone, *Health of the people: the highest law?* in Dawson 2007, p. 59-77

Hakulinen 2011	T. Hakulinen et al, 'Harmonization may be counterproductive – at least for parts of Europe where public health research operates effectively', letter to *EJPH* (accepted for publication)
Hansson 2009	M.G. Hansson, 'Ethics and Biobanks', *BJC* 2009, 100, p. 8-12
Harris 2005	J. Harris, 'Scientific research is a moral duty', *J Med Ethics* 2005 31:242-248
Heller 1998	M.A. Heller & R.S. Eisenberg, 'Can patents deter innovation? The anticommons in biomedical research', *Science* 1998, 280(5364), p 698-701
Holm 2011	S. Holm, 'Systems, rules and the costs of being ethical – a response to D. Chalmers and to S. Whitney and C. Schneider' *Journal of Internal Medicine* 2011, DOI 10.1111/j.1365-2796.2011.02370.x, p. 403-406
Homer 2008	N. Homer et al, 'Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays', *PLoS Genet*, 4(8): e1000167. doi:10.1371/journal.pgen.1000167
Herk-Sukel 2010	M.P.P. van Herk-Sukkel et al, New Opportunities for drug outcomes research in cancer patients. The linkage of the Eindhoven Cancer Registry and the PHARMO record linkage system, *EJC* 2010, 46, p. 395-404
Ingelfinger 2004	J.R. Ingelfinger, J.M. Drazen, 'Registry research and medical privacy', *NEJM* 2004, 350 (14), p. 1452-1453
ISO 2008	International Organization for Standardization, *Health informatics – Pseudonymization ISO/TS 25237:2008*, Geneva: ISO, 2008
Jacobsen 1999	S.J. Jacobsen et al, 'Potential effect of authorization bias on medical record research', *Mayo Clinic Proc* 1999, 74 (4), p. 330-338
Jennings 2007	B. Jennings, *Public health and civic republicanism: toward an alternative framework for public health ethics*, included in Dawson 2007
Jousilahti 2005	P. Jousilahti et al, 'Total and cause specific mortality among participants and non-participants of population based surveys: a comprehensive follow-up of 54372 Finnish men and women', *J Epedimiol. Community Health* 2005, 59, p. 310-315
Knoops 1999	B.J. Knoops, 'De Trusted Third Party bestaat niet', *Informatie ??* 1999, 41, p. 40-41
Knoppers 2005	B.M. Knoppers & R. Chadwick, 'Human genetic research: emerging trends in ethics', *Nat Rev Genet* 2005, 6 (1), p. 75-79
Korff 2010	D. Korff, *Data protection laws in the EU: The difficulties in meeting the challenges posed by global social and technical developments*, Brussels: European Commission Directorate-General Justice, Freedom and Security (Working Paper No. 2, part of a Comparative Study on different approaches to new privacy challenges, in particular in the light of technological developments) 2010
Kühn 2004	H.C. Kühn, *The Implementation of the Data Protection Directive 95/46/EC in Germany*, included in D. Beyleveld et al (ed), *Implementation of the Data Protection Directive in Relation to Medical Research in Europe*, Aldershot/ Burlington: Ashgate 2004
Legemaate 2004	J. Legemaate, 'Het zelfbeschikkingsrecht: een oud debat in nieuw licht', *TvGR* 2004, 1(28), p. 18-20
Lenaerts 2011	K. Lenaerts & P. van Nuffel, *European Union Law*, London: Sweet & Maxwell 2011
Lowrance 2002	W.W. Lowrance, Learning from Experience, Privacy and the secondary use of data for health research, London: Nuffield 2002
Lowrance 2003	W.W. Lowrance, Learning from experience: privacy and the secondary use of data in health research. J. health Serv. Res. Policy, Vol. 8, supl. 1 July 2003, p; 2-17, op p. 3-4.
Lowrance 2007	W.W. Lowrance, F.S. Collins, 'Identifiability in Genomic Reseach', *Science* 2007, p. 600-603

Macklin 1993 R. Macklin, *Enemies of patients: how doctors are losing their power & patients are losing their rights*, New York/Oxford: Oxford University Press 1993

Mancini 2010 J. Mancini et al, 'Consent for Biobanking: assessing the understanding and views of cancer petients', *JNCI* 2011;103:1-4

Mascalzoni 2008 D. Mascalzoni et al, 'Informed consent in the Genomics Era', *PLoS Medicine* 2008, 5 (9), p. 1302-1305

Mooij 2004 A. Mooij, *Geen paniek!, Aids in Nederland 1982-2004*, Amsterdam: Bert Bakker 2004

Mordden 1997 E. Mordden, *How long has this been going on?* Stonewall Inn Editions 1997

Newdick 2011 C. Newdick, 'Disrupting the community – saving public health ethics from the EU internal market', in van de Gronden 2011

Nissenbaum 2010 H. Nissenbaum, *Privacy in context: technology, policy and the integrity of social life*, Stanford California: Stanford University Press 2010

O' Neill 2004 O. O'Neill, 'Accountability, trust and informed consent in medical practice and research', *Clinical Medicine* 2004, 4 (3), p. 269-276

Ohmn 2010 P. Ohm, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization (August 13, 2009). *UCLA Law Review* 2010, Vol. 57, p. 1701; U of Colorado Law Legal Studies Research Paper No. 9-12. Available at SSRN: http://ssrn.com/abstract=1450006

Olsthoorn-Heim 2003 E.T.M. Olsthoorn-Heim, Het dossier lichaamsmateriaal, *Tijdschrift voor Gezondheidsrecht* 2003, 5, p. 306-313

Ploem 2004 M.C. Ploem, *Regulering van gegevensverwerking voor medisch-wetenschappelijk onderzoek (diss. Amsterdam UvA)*, Amsterdam: 2004

Ploem 2010 M.C. Ploem, *Gegeven voor de wetenschap; Regulering van onderzoek met gegevens, lichaamsmateriaal en biobanken*, in Wetenschappelijk onderzoek in de zorg: Preadvies 2010, Vereniging voor Gezondheidsrecht, Den Haag, SDU 2010, p. 117-210

Plomer 2005 A. Plomer, *The Law and Ethics of Medical Research: International Bioethics and Human Rights*, London/Sydney/Portland: Cavendish Publishing 2005

Porter 1997 R. Porter, *The Greatest Benefit to Mankind: a medical history of humanity from antiquity to the present*, London: HarperCollins Publishers 1997

Poulson 1998 J. Poulson, 'Bitter pills to swallow', *N Engl J Med* 1998; 338:1844-1846

Privireal 2005 *Recommendadtions from PRIVIREAL to the European Commission*, as published by PRIVIREAL (Privacy in Research Ethics & Law, an EC Framework 5 project) on www.privireal.org)

Quathem 2005 K. van Quatherm, 'Controlling personal data – the case of clinical trials', *P&I* 2005, p. 155-161

Registratiekamer 1995/2000 H. van Rossum e.a., Privacy-enhancing technologies: the path to anonymity, volume I and II. A&V 5; Registratiekamer, Den Haag 1995, revised edition 2000

Reng 2003 C.M. Reng et al, 'Akzeptiertes Datenschutzkonzept', *Deutsches Ärztenblatt* 2003, 100(3)

RGO 2008 Raad voor Gezondheidsonderzoek, *Van gegevens verzekerd. Kennis over de volksgezondheid in Nederland nu en in de toekomst*, Den Haag: Gezondheidsraad, 2008, RGO nr 58

Riegman 2011 P. Riegman, E.B. van Veen, Research with residual tissue, *J. Human Genetics* 2011, 130 (8), DOI: 10.1007/s00439-011-1074-x

Rose 2007 N. Rose, *The Politics of Life Itself: biomedicine, power and subjectivity in the twenty-first century*, Princeton/Oxford: Princeton Un. Press 2007

Sen 2009 A. Sen, *The Idea of Justice*, London: PenguinBooks 2010 (Allen Lane 2009)

Shapshay 2007 S. Shapshay & K.D. Pimple, 'Participation in biomedical research is an imperfect moral duty: a response to John Harris', *JME* 2007, 33: 414-417

Stolker 2003 C.J.J.M. Stolker, 'Ja, geléérd zijn jullie wel!; over de status van de rechtswetenschap', *NJB* 2003, 15, p. 766-778

Sunstein 2005 C.R. Sunstein, *Laws of Fear: beyond the precautionary principle*, Cambridge/New York: Cambridge University Press 2005

Sutrop 2011 M. Sutrop, 'How to avoid a dichotomy between autonomy and beneficence: from liberalism to communitarianism and beyond', *Journal of Internal Medicine* 2011, DOI 10.1111/j.1365-2796.2011.02370.x, p. 375-379

Tambor 2010 M. Tambor et al, 'Diversity and dynamics of patient cost-sharing for physicians' and hospital services in the 27 European Union countries', *European Journal of Public Health* 2010, 1-6

Ten Haaft 2010 G ten Haaft, *Dokter is ziek. Als patiënt zie je hoe de zorg beter kan*. Amsterdam/Antwerpen: Uitgeverij contact, 2010

The 2010 A.M. The, *Tussen hoop en vrees*, Bohn Staflau van Loghum 2010

Tu 2004 J.V. Tu et al, 'Impracticability of informed consent in the registry of the Canadian Stroke Network', *N Engl J Med* 2004; 350, p. 1414-1421

Van de Gronden 2011 J.W. van de Gronden et al (editors), *Health Care and EU Law*, The Hague: T.M.C. Asser Press 2011

Van der Valk 2010 T. van der Valk, 'Zeggenschap over lichaamsmateriaal: suggesties vanuit Europa', *TvGR* 2010, 7, p. 562

Van Hurk-Sukel 2010 M.P.P. van Hurk-Sukel et al, 'New opportunies for drug outcomes research in cancer patients; the linkage of the Eindhoven Cancer Registry and the PHARMO Record linkage system', *EJC* 2010, p. 395-404

Van Leeuwen 2008 E. van Leeuwen, 'Ethiek van medisch-wetenschappelijk onderzoek: informed consent en de therapeutische misconceptie', *NTvG* 2008, 12, p. 679-683

Van Veen 1994 E.B. van Veen, J.M. Buiting en M.A. Goslings, 'Codering kan beroepsgeheim bij medisch onderzoek waarborgen', *NRC Handelsblad* 1994, artikel op 8 januari.

Van Veen 2003 E.B. van Veen, 'Gecodeerde gegevens bij wetenschappelijk onderzoek: een begripsverheldering', *P&I* 2003, December, p. 259-262

Van Veen 2005 E.B. van Veen & G.J. Hamilton, 'De Zorgverzekeringswet in Europees(rechtelijk) perspectief', *TvGR* 2005, 1, p. 80-97

Van Veen 2006 E.B. van Veen et al, 'TuBaFrost 3: Regulatory and ethical issues on the exchange of residual tissue for research across Europe', *EJC* 2006, 42 (17), p. 2914-2923

Van Veen 2008 E.B. van Veen, 'Obstacles to European research projects with data and tissue: solutions and further challenges', *EJC* 2008, 44 (10), p. 1438-1450

Van Veen 2008b E.B. van Veen, Boekrecensie (bespreking van: J.A. Bovenberg, *Property Rights in Blood, Genes and Data: naturally yours?*, Leiden/Boston: Martinus Nijhoff 2006), *TvGR* 2008, 1, p. 75-78

Van Veen 2008c E.B. van Veen & I. Janssen, *Toetsing op Maat*, Den Haag: MedLawconsult 2008

Vandenbroucke 2010 J.P. Vandenbroucke, 'Preregistration of Epideniologic Studies – an ill-founded mix of ideas', *Epidemiology* 2010, 21 (5), p. 619-620

Veach 2009 R.M. Veach, *Patient, Heal Thysel: how the new medicine puts the patient in chargef*, Oxford/New York: Oxford University Press, 2009

Verschuuren 2008 M. Verschuuren et al, 'The European data protection legislation and its consequences for public health monitoring: a plea for action', *European Journal of Public Health* 2008, 1-2

Walzer 1983 M. Walzer, *Spheres of Justice: a Defence of Pluralism*. New York 1983: Basic Books

WHO 2005 World Health Organisation, *International Health Regulations (2005)*, WHO 2005

Wilson 1968 J.M.G. Wilson & G. Junger, *Principles and practice of screening for desease*, Geneva: World Health Organisation 1968

Woolf 2000	S.H. Woolf et al, 'Selection bias from requiring patients to give consent to examine data for health services research', *Arch Fam Med* 2000, 9 (10), p. 1111-1118
Working Party 2007a	*Working Document on the processing of personal data relating to health in electronic health records*, Article 29 Data Protection Working Party, Brussels, 2007: 00323/07/EN, WP 131
Working Party 2007b	*Opinion 4/2007 on the concept of personal data*, Article 29 Data Protection Working Party, Brussels 2007: 0124807/EN, WP 136
Working Party 2007c	Opinion 2/2007 on information to passengers about transfer of PNR data to US *Authorities,* Article 29 Data Protection Working Party, Brussels 2007: 000345/07/EN,WP132
Working Party 2010	*Opinion 1/2010 on the concepts of "controller" and "processor"*, Article 29 Data Protection Working Party, Brussels, 2010: 00264/EN, WP 169
WRR 2000	Wetenschappelijke Raad voor het Regeringsbeleid, *Het borgen van publiek belang (r56 2000)*, Den Haag: Sdu 2000

About the author:

Evert-Ben van Veen Ll.M. has been active as a Legal consultant in health care for many years. Before he started with MedLawconsult he was amongst other things head legal counsel of the University Hospital Rotterdam (at present Erasmus MC) and of the National Council for Public Health (now defunct). With MedLawconsult he has been involved in various European projects, either as a partner or as a subcontractor, such as TubaFrost, VIRGIL, ENTB-GENEPI, TRANSFoRm and at present EUROCOURSE.
He is both interested in the more fundamental questions of law, ethics and health care and the practical applications of law in health care, coming down to such things as contracts which work for all parties concerned (the process of getting there being just as important as the signed agreement) and counsel to clients, concrete and down to earth. Such contracts and counsel always have to adhere to existing law of course. In his more fundamental works Evert-Ben van Veen likes at times to challenge existing views. The present report has elements of both.

A list of publications can be found under ‘publicaties’ at www.medlaw.nl

This Report was made possible with financial support from:

FEDERA 

COREON 

EUROCOURSE 

MedLawconsult is a small legal consultancy firm, based in The Hague, the Netherlands. Clients range from small e-health start ups to governmental agencies and international research consortia. Most of our clients are not for profit organisations. Our aim is to make a meaningful contribution, both to our clients and to health care in general. We have a thorough knowledge of Dutch law, of European law (limited to fields relevant to our clients) and international regulations concerning medical research, both clinical and observational.

ISBN 978-90-75941-00-5

9 789075 941005

MedLawconsult, Postbus 11500, 2502 AM Den Haag. Tel. 070-3589772