Chapter 9: The Principle of Autonomy: Privacy and Consent in a world of Big Data and Genetic Data

The principle of *Autonomy* is that researchers and scientists have a duty to respect the ability of the patient or research participant to make their own informed decisions. In research projects this is typically taken to imply (a) a need for informed consent and (b) an obligation to protect the privacy of patients or research participants. But, in a world where scientific jargon is often impenetrable to non-experts, what does *informed* consent mean, in practice? And what about vulnerable populations, who many not feel able to withhold consent?

Since Autonomy must be balanced with other principles, there may be times when 'consent' is not reasonable if this restriction may cause harm to others – this is relevant in the case of HeLa cells, used without consent in the development of medicines and in research. Life science researchers are also faced with multiple threats to *privacy* – whether through problems of data security, through the power of big data analysis tools to make identifying deductions about people, or through the capacity to extrapolate from the genetic material of one individual to that person's relatives.

How informed?: Non-steroidal anti-inflammatory use by rugby players

Rugby is a contact sport which involves heavy physical impacts between players who typically range of 90kg to 120kg. Players will often experience pain after matches even without injury, and muscle and ligament injuries are relatively frequent.

Paul O'Connell is one of the all time great rugby players, who won 2 European cups for his club (Munster) and 3 six nations championships. By 2010 he had been playing for a decade at the highest level, had accumulated multiple injuries, and at the same time, was made captain of the Irish rugby team. He later wrote about his use of non-steroidal anti-inflammatories such as Diclofenac and Ibuprofen to allow him manage pain enough to be able to play: "In that little battle with my own body, anti-inflammatories were my friend. After taking Difene [Diclofenac], I felt like I could play for another five years….Towards the end of my career, for a Saturday game I took Difene on Friday and Saturday. It was like a miracle drug, except it could disagree with my stomach. So I took Zoton, an indigestion tablet that dissolves on the tongue and lines the stomach. By kick-off the body felt – at least for a while – pretty much

symptom-free. No aches, no pains, nothing. I was always conscious that anti-inflammatories needed to be taken in moderation. I hated it whenever I had to take Difene to play".

Lewis Moody is an English rugby player who played at the same time as Paul O'Connell, winning 2 European Cups for his club Leicester (one against Paul O'Connell's Munster), 3 six nations championships, and a Rugby World Cup for England. Moody has said: "I was taking drugs so I could play, like ibuprofen and diclofenac. It was like I was a walking medicine cabinet. I don't think I'd change much about my life, but I would probably change my lax approach to this…I remember one story. We were on a bus. It was almost like a kind of challenge to see how many 'smarties' we could take... I didn't ask questions then. I don't think you'll ever change the single-minded sportsman, but I think they could be better informed". Lewis Moody was diagnosed in 2005 with ulcerative colitis a chronic (incurable) disease which lead to pain, blood in stools, severe diarrhoea and weight loss.

The information sheet for Ibuprofen includes identifies that digestive tract problems are a rare side effect: (1 in 10,000 people) "may effect stomach or intestinal ulcers, sometimes with bleeding and perforation, inflammation of the lining of the mouth with ulceration (ulcerative stomatitis), inflammation of the stomach (gastritis)". It identifies that there is a risk of worsening colitis and Crohn's disease, but the risk is not quantifiable. The Diclofenac information sheet identifies that there is a 1 in 10,000 risk of "inflammation of colon causes abdominal pain and diarrhoea, aggravating of existing digestive problems such as colitis or Chron's disease".

1. The care concept of 'caring for' (responsibility) suggests we need to locate our actions in a social network with empathy for others. The people involved in this social network

include the (a) player, (b) the company that makes and supplies the medicines, (c) the player's teammates, and (d) the players' (or their teams') doctors. Are there other important people in the social network you would mention?

2. Let's focus on the (a) player, (b) the company that makes and supplies the medicines, (c) the player's teammates, and (d) the players' (or their teams') doctors. For each of these, identify their perspective on the experience described here.
3. For each of the four, identify what emotions they would probably have felt at the time. Identify what emotions they would probably feel now. What are the thought action tendencies associated with each of these emotions (refer back to chapter 3 if needed)?
4. Based upon what you know of 'informed consent', and taking into account the idea that we should try to construct solutions which take care of all those who need and give care in such as situation, would you suggest any changes to the way people are informed about medicine use in such cases?
5. The concept of 'receiving care' (responsiveness) identifies that we need monitor how the care we are proposing is being received. How might that be applied in this case?

Introduction

The focus of this chapter is on issues related to the principle of 'autonomy' in life sciences engineering and research ethics.

As you saw in chapter 6, one of the four principles of bioethics as it developed in the post-war period was the principle of Autonomy or Respect for Persons*:* The core idea in the principle of autonomy is that each person should be free to make their own decisions about things which are important to them. This idea was seen as being akin to applying the principle of democratic self-rule to an individual:

> . . . the core idea of personal autonomy is an extension of political self-rule to self-governance by the individual: personal rule of the self while remaining free from both controlling interferences by others and personal limitations such as inadequate understanding, that prevent meaningful choice (Beauchamp and Childress 1989: 68).

In the case of research projects, this principle was translated, in practice, into the idea of 'informed consent' which, in turn was written into numerous codes of ethics. As you saw in chapter 7 this is written into the Declaration on Human Rights and Bioethics which states that any "preventative, diagnostic or therapeutic medical intervention", and any "scientific research", should only be "carried out with prior, free and informed consent based on adequate information" (Article 6). It is also fundamental to other codes of ethical practice such as the Nurenburg Code and the Declaration of Helsinki. As you saw in chapter 8, it is also written into Swiss law. Article 7 of the Federal Act on Research involving Human Beings (2011) states "Research involving human beings may only be carried out if, in accordance with the provisions of this Act, the persons concerned have given their informed consent or, after being duly informed, have not exercised their right to dissent. The persons concerned may withhold or revoke their consent at any time, without stating their reasons".

Although these various documents, codes and laws seems to enshrine an clear requirement for 'informed consent', in reality, informed consent can be understood as existing on a continuum between contexts in which it is required and its form is strictly controlled, through circumstances when it required but less strictly controlled, to circumstances where consent is implied.

'Normal' informed consent

The Human Research Act Article 16 and 17 states: Persons may only be involved in a research project if they have given their informed consent.. Consent must be given in writing; the Federal Council may specify exemptions.

The persons concerned must receive comprehensible oral and written information on:

a. the nature, purpose and duration of, and procedure for, the research project;
b. the foreseeable risks and burdens;
c. the expected benefits of the research project, in particular for themselves or for other people;
d. the measures taken to protect the personal data collected;
e. their rights.

Before a decision on consent is made by the persons concerned, they must be allowed an appropriate period for reflection.

If the intention exists to make further use for research of biological material sampled or health-related personal data collected, the consent of the persons concerned must be obtained at the time of such sampling or collection, or they must be informed of their right to dissent.

Questions:

1. The law requires consent in writing (this is normally taken to mean signed by the research participant). How do you think this may appear from their perspective? What emotions might they experience when asked to sign an informed consent form?
2. It is required that the information includes the measures to protect data security. Why do you think data security is felt to be so important that it explicitly mentioned in the Act? Under normal circumstances, in your everyday life, how much attention do you pay to your data security?

This can be taken as the 'normal' informed consent procedure (orally informed and in writing, with some delay for reflection, and with a person signing a consent form that contains certain legally required information). Implicit in this is the idea that the signing of informed consent is to protect the patient or subject (i.e. it ensures they are fully informed before making a decision), and that it is a moment of decision (i.e. they have time to reflect in order to arrive at a decision).

From an ethics of care perspective, we want to be able to empathise with the people involved in a process and so it is valuable to look to social research to find out how the research participants/ patients understand this process from their perspective. Research has been undertaken with people who have taken part in medical research to see how they understand and experience the informed consent process. Corrigan (2003) identified that, while formal ethics processes intend informed consent understand the informed consent process as being a moment of decision and a protection of the research subject, participants often understood the process differently. She identified:

- Participants often don't see informed consent as representing a real decision that they make
- Participants don't adequately understand risks
- Participants don't understand the research process they are going through

Is informed consent a real moment of decision?: For Corrigan, many of the participants had already made up their mind to participate in the research before getting to the stage of informed consent (Lewis and Graham, 2007 had similar findings). Sometimes this was linked to a sense of trust in the medical professionals involved. As one research participant in a Phase I trial (which aimed at testing whether a treatment was safe) stated:

> I thought if there was anything wrong or anything that could be allowed to happen that would cause long term injury or anything, then they wouldn't be allowed to do these sorts of studies. I thought it has got to be pretty safe for them to be allowed to do it.

Another patient in a clinical trial said in interview:

> I saw the doctor and she said would I like to go in for this new drug, and I said 'I don't know anything about it, it's up to you, if you think it will do me good, all right I will go on it'

A second reason why people had already decided to participate was on the expectation that the study would involve them getting 'better' care than if they did not participate – it was widely assumed that the reason a new treatment was being tested was that there were strong reasons to believe it was better than existing treatments. Implicit in this was a misunderstanding of the success rates of clinical trials: participants generally assumed that most trials are successful (in reality only 25% - 30% of treatment trials will pass Phases I, II and III). Some of those who participated expressed dissatisfaction with their existing treatment which they implicitly saw as being inadequate. Prostate cancers are often slow to develop and since treatments can have worse outcomes than no treatment, 'watchful waiting' (in which a person is monitored regularly to see if the disease is progressing), is often the most appropriate treatment. For the patient in this case the idea that regular monitoring (which they experience as being 'no treatment') is the best treatment may be hard to accept. Hence Corrigan found they were disposed to participate in research since any treatment was perceived as better than 'watchful waiting'.

How informed are people about risks?: Multiple researchers have found that, after going through an informed consent process, participants were unable to subsequently recall any of the side-effects listed (Bergler *et al.* 1980; Estey *et al.* 1994). Hence there are questions as to how 'informed' the process really is.

How informed are people about the research process?: Participants know something about the research process before they become a research participant, but they sometimes don't know a lot. In chapter two we met various cognitive biases that impact on people's decision-making. One such bias is *confirmation bias*, in which people interpret new information in light of their prior beliefs. This was evident in Corrigan's research in that some of her respondents interpreted the research study in terms of their prior beliefs and understandings of research studies and therefore did not notice or understand key aspects of the research. In one case, a research participant who had agreed to participate had a prior assumption that all clinical trials involve a placebo. In their case, the trail involved comparison of a new treatment with a standard treatment, however their prior beliefs led them to not recognise this even when they accurately recalled the information on the information sheet:

> There are two sorts that one can take . . . but whether . . . which one I am on I am not sure. Which one I am on I haven't a clue, and *whether I am not on any I am not sure*. It didn't say in the thing [information sheet] that I might be given nothing but I don't know (emphasis added).

In another interview, a study (with a somewhat complicated trial structure which involved comparing four conditions, one of which was a placebo), a participant had the following exchange with Corrigan (marked as OC in the transcript):

> OC: Do you know what your chances are of receiving the trial drug?
> Patient: I know for certain I am on a drug but I don't know what it is because they haven't told me.
> OC: But are you definitely on some form of drug or medication?
> Patient: Yes, yes.
> OC: What do you know about the different drugs being tested?
> Patient: Well I know there are two, but they don't tell you which you are on, which is fair enough.

A second source of cognitive bias which can impact on people's understanding is the wording used in consent forms. Words which are used as synonyms may in fact have quite different implicit meanings to people. For example, a study on patients' attitudes towards a cancer drug trial found that 72% found the prospect appealing when it was said that the trial was for a

'new' treatment, but only 27% thought that the prospect was appealing when the treatment was described as 'experimental' (Slevin et al., 1995). Similarly, Corrigan (2003) suggests people may be more likely to participate in something described as a 'study' than if the same trial is described as an 'experiment'.

Informed consent as induced compliance: It is not just words that have implicit meanings for people, but also acts. In many cultures the act of signing a document establishes the idea that the person is making a commitment. As such, participants may not see signing an informed consent form as being only an acknowledgement that they have been informed – rather they may see it as making a commitment to participate in the study. Swan and Collins (2008) tested this idea by running an experiment in which half of the participants had the information about the study presented to them orally and in writing, while the other half were given the same information in the same way but were also asked to sign the information sheet. Participants were then told that there was a problem with the experiment and were asked if they could come back one week later. Those who had signed were more likely to agree to come back later than those who had not signed (79.5% compared to 59.1%). In a different experiment, those who had signed were more likely to stay in an experiment that took longer than anticipated when compared to those who had not signed. The researchers concluded:

> Our research indicates that informed consent protocols …as the means to inform participants of their rights may, in fact, be decreasing the probability that they will function as autonomous agents and exercise those rights… Rather than informing participants of their rights, our data indicate that signing the [Informed Consent Form] creates a commitment to comply (2008, 2642-2643).

Question

1. It appears that many participants do not see the informed consent process in the way that it is envisaged in the legal and regulatory frameworks. Is this an ethical problem? Why, or why not?

2. Let's focus on the (a) research subject, (b) the researcher, (c) other patients who suffer from an illness that may be treated by the treatment being tested, and (d) the wider public in a Phase II trial (i.e. a treatment that has been found to have acceptable risks but for which effectiveness is not yet determined). For each of these, identify their perspective on someone being asked to participate, under the current informed consent process.
3. For each of the four, identify what emotions they would probably feel in relation to a person being asked to give informed consent. How might those emotions impact on their thinking and acting?
4. Taking into account the idea that we should try to construct solutions which take care of all those who need and give care in such as situation, would you suggest any changes to the way people are informed about trials?
5. Taking into account that we should be responsive to how people experience the care provided, what follow up processes would you recommend?

Restricted informed consent

In the case of particularly vulnerable people, there are additional restrictions on informed consent. Particularly vulnerable people are taken to mean (under Swiss law):

- Children (which means people under the age of 14 in Switzerland)
- Adolescents (14 to 18 years old)
- Adults "lacking capacity"

In this case, a legal representative gives consent. The person themselves is to be involved in consent if they are "capable of judgement". If they are identified as lacking capacity, then consent is presumed not to exist if they " visibly express opposition to the research intervention either verbally or by his or her behaviour".

Other vulnerable people (pregnant women and prisoners) have restrictions in place on what research can take place, but normal informed consent rules apply.

Question

1. Medical students (or life sciences engineering students) and hospital staff may be under the control of a person running a research study (e.g. pays their salary or controls their grades). Are there circumstances in which they might they feel compelled/pressured to participate in a study? Should they be subject to particular protections?
2. Are there other particularly vulnerable groups that you would think should be protected?

Implied consent

There are some circumstances in which a person is assumed to consent unless they say otherwise. In Swiss law there are essentially, two kinds of contexts in which this applies:

- Emergency contexts, where consent may not be possible due to a person's medical condition and the lack of time to engage with family or legal guardians etc., and
- Times when access to data is deemed more important than an absolute right to privacy with respect to health or biological data.

In emergency procedures, consent can be assumed if (a) you try to determine their consent as soon as possible, (b) the person does not visibly express opposition to the research intervention through either verbally or by his or her behaviour, (c) a doctor not involved in the research project safeguards their interests. In this case, post hoc consent is sought from the person as soon as possible.

The second context in which consent may not be needed refers to the secondary use of data for research purposes, i.e., when the person is not being asked to consent to getting an experimental treatment but where their data is being used in research. In such cases, explicit consent *is* required for the re-use of biological material or health data if the person involved is identified or identifiable. Explicit consent *is not* required in the case of 'anonymised' health material used for research purposes. This includes:

- Anonymized biological material and genetic material
- De-identified non-genetic health-related personal data ('de-identified' means that there is no identifying data included but a code exists which can link the data back to a person's identity if needed)

In some cases, for identified or identifiable material may not require informed consent. identified or identifiable. Further use may be made of biological material or health-related personal data for research purposes in exceptional cases if:

a. it is impossible or disproportionately difficult to obtain consent or to provide information on the right to dissent, or this would impose an undue burden on the person concerned;
b. no documented refusal is available; and
c. the interests of research outweigh the interests of the person concerned in deciding on the further use of his or her biological material and data.

Consent related to personal health information

Privacy rights and HeLa Cells
HeLa cells are human cells that reproduce continuously and can therefore be grown for prolonged periods in vitro. HeLa cells are the oldest human cell line and one of the most

commonly used. In medicine, HeLa cells have played important roles in the eradication of Polio, in development of cancer treatment, and in mapping the human genome. They are also used in education (you may well have used HeLa cells yourself in your coursework).

The name "HeLa" refers to the name of a woman, Henrietta Lacks, who was, in 1951 a 30-year old mother of five children living in Baltimore. She had grown up working as a child on a tobacco farm in rural Virginia before moving to Baltimore with her family. Not long after the birth of her fifth child, she was admitted to hospital for treatment for vaginal bleeding which led to a diagnosis of a rare and aggressive form of cancer. Biopsies from her cervix were taken and were sent to the Tisue Culture Laboratory to be studied. It was not normal practice at that time to ask for written permission to obtain such samples for research purposes, and there is no record that Henrietta Lacks consented to the use of her cells. Unlike previous human cell lines, these cells did not die out but were rapidly propagated. Henrietta Lacks died in October 1951, leaving her husband David, and her five children Lawrence, Elsie, David Deborah and Joseph (who was less than one-year old when she died).

Henrietta Lacks' tissue samples were developed and became widely used in scientific research. Her cells were central to the development of vaccines for Polio, HPV, and COVID, to cancer research and to the development of in-vitro fertilisation. Writing in Nature Cancer Reviews, John R. Masters said:

> Our knowledge of every fundamental process that occurs in human cells — whether normal or abnormal — has depended to a large extent on using HeLa and other cell lines as a model system. Much of what we know today, and much of what we do tomorrow, depends on the supply of HeLa and other cell lines (2002: 316)

Alongside important medical and scientific discoveries was a significant breech of her privacy:

…for decades after her death, doctors and scientists repeatedly failed to ask her family for consent as they revealed Lacks's name publicly, gave her medical records to the media, and even published her cells' genome online (Nature, 2020: 7).

Alongside the privacy issues, biotechnology companies made significant income from the exploitation of Henrietta Lacks' – HeLa – cells. Her family did not become aware of the use of her cells for decades after her death. In 2021 they filed a lawsuit against Thermo Fisher Scientific – a biotechnology company that uses her cells. The case was settled in 2023 without the terms of the settlement being announced (CBS News, 2023).

Questions:

1. Henrietta Lacks' cells have contributed to enormous scientific and medical benefits which may not have developed as quickly or at all had her cells not been available or had she refused consent for their use. How do you balance her autonomy (right to informed choice as to the use of her cells) with the social and health benefits of the use of her cells?
2. In Henrietta Lacks' case, her actual tissue was used. Imagine if only her medical records (i.e. no biological material) was needed to achieve such benefits. Would her right to privacy (right to withhold access to her data) outweigh the social benefit?

Consent is generally framed in terms of people agreeing to the physiological or bio-health risks related to a study or intervention. But, as you can see in relation to implied consent in Swiss law, sometimes the consent relates to a person's data or the privacy of their biological or health related data.

The availability of health related data for secondary research is increasingly important in the context of the growing availability of 'big data' research techniques. 'Traditional' research involved a researcher with a hypothesis identifying and collecting the data needed to test the

hypothesis. In order to protect the privacy of participants, part of this approach was to only collect data that needed to be collected to test the specific hypothesis and then destroying data when it was no longer needed. This approach, to privacy, called 'data minimisation' , was consistent with traditional research and is also built into data protection legislation. Since the development of 'big data' techniques, new modes of research have developed in which researchers find unexpected patterns in data which are predictive of particular outcomes. This means finding patterns in large datasets which were not previously hypothesised by researchers but which nonetheless can tell something meaningful. For example, the standard test today to determine risk of prostate cancer is a blood test for Prostate Specific Antigen (PSA). However elevated PSA only relates to an enlarged prostate, not to cancer as a cause of the enlargement. Hence elevated PSA gives a false positive for cancer in 75% of cases. A recently published study on the use of 130 genetic markers to predict prostate cancer was found to provide accurate predictions 40% of the time (Devlin, 2025). The test is based on saliva and therefore is also far less intrusive and easier to administer than the PSA test. This kind of breakthrough is only possible due to the capacity to process large datasets with sufficient data to allow for predictive models for health outcomes to be developed and tests. This kind of situation – in which we need to have large datasets with lots of variables which are not driven by a specific hypothesis – seems at odds with the approach to data minimisation approach to privacy which developed in the 1960s and 1970s and which was adapted to 'traditional' experimental research.

In Swiss law, part of the context in which genetic data could be used for secondary research was that it was *anonymous*. It is worthwhile to explore that term in more depth. Anonymity is not as simple as 'not being identified' – the term 'anonymity' implies that a person is neither *identified* nor *identifiable*. The term 'identified' is probably straightforward enough, it

means that the data does not have any identifiers (e.g., name, social security number, student id card number, your mobile phone IP address etc.) linked to a specific person. However it is possible that data may have no identifiers and still allow a person to be identified. In research, it may be necessary to, at some stage in the future, trace an individual. In this case the researcher will often remove identifiers and add a unique code, then, separately, keep a table of the identity which goes with each code. Someone who has both the data and the code table can identify individuals, but if someone has only the dataset then they will not be able to 'connect' back to the original identifier. This is called 'de-identified' data, and is acceptable in the Swiss law for non-genetic health-related personal data.

It is sometimes possible for deidentified data to still allow a person to be identified. Imagine someone had collected data on age, and PSA (prostate specific antigen) from the participants who attend our classroom on some Thursday at 16h15. With three pieces of information (age, present in the classroom, has a prostate), someone would be able to identify health related information about me (since I am the only person in the classroom who is over 50 and has a prostate). In this case, the data may have no identifier (name, social security number etc.) but I may still be identifiable.

In this way, if there are sufficient different variables gathered about a person in a dataset, then a person in the dataset may be identified by crossing different variables. In the case above the set which contained 'people in the class', 'people with a prostate' and 'people over 50' has a single person (me). In general, the more variables which are collected about people in a dataset, the easier it is to find a unique combination of variables. Hence, in traditional studies, it was normally expected that researchers will only collect data on a given variable if they have a reason to do so (this is part of what is referred to as 'data minimisation'). However

this is only appropriate from a research perspective in a traditional study which has been designed to test a specific hypothesis.

A second case may arise with a dataset which, in itself, may not allow a person to be identified to a stranger, but something in the dataset may make someone identifiable to someone who has additional information about them. The person with 'additional information' may be someone who knows them, or, in a world of social media, may simply be someone who has access to the internet.

For example, Saunders, Kitzinger and Kitzinger (2015) describe the challenges in offering anonymity to research participants in a study of the experiences of family members of people with severe brain injuries. In the interviews, participants shared a great deal of sensitive information including their perceptions about how other family members responded to the injured family member (which could cause family conflict if it became known), how they perceived doctors (which they felt might affect their family member's care if it became known) and their perception of whether or not the family member would have wanted to be kept alive were they able to express their wishes (which could cause them guilt, or shame and could anger others if it became known).

While the researchers could ensure that the quotes they included in their research reports did not allow individuals to be identified, they could not ensure that information that was in the reports could not be connected to data from other sources which would, taken together, allow the person to be identified. In particular, where family members had shared details on social media (e.g., Facebook etc.), in media interviews or in court cases related to the family member's injury, these publicly available sources of data, when combined with the interview

quotes could allow a research participant and their sensitive data to be identified (this is referred to as 'deductive disclosure'). The researchers identified that there was no single answer to this problem:

> attempts to address anonymizing issues were made through collaboration with interviewees and reflecting with them on how important anonymity could be for them and whether some of the material they shared with us was more sensitive than others. We found a range of views among interviewees about the level of protection they wanted. Some interviewees felt we were being overprotective and even resisted efforts to ensure their anonymity—this could include, for example, an express wish that we use not only their real name but also the real name of the patient (something we could not usually do as the patient was unable to give consent). Other interviewees wanted to maximize anonymity including, for example, ensuring that some of what they said would not be recognized by other family members (2015: 131-132).

Informed consent when anonymity is impossible

This is an extract of the informed consent form used by Saunders, Kitzinger and Kitzinger (2015) to explain the process of de-identification and the impossibility of anonymity:

We will try to ensure that nobody can identify you from extracts from your interview in the following ways:
1. We will change your name and the names of people in your family, other personal contacts, and the names of the professionals who cared for your loved one.
2. We will change the names of any hospitals, residential care homes, or rehabilitation units that you mention—and the names of towns and cities where they are located and/or where you live.
3. Unless you specifically give consent to the contrary, we will modify your occupation (if you mention it) and that of the person with brain injury to make you less identifiable.
4. We will alter or remove any other details you request. We can also remove some extracts from your interview and assign them a different pseudonym and identification number so that they cannot be identified as having been spoken by the same person.

However carefully we anonymize your interview…it might be possible for someone …to identify you when they read extracts from the interview you did with us. For example, this might happen if you have an unusual story with distinctive features that people will recognize, if you use very similar words and phrases …

We hope you understand that if you are "going public" in other contexts we cannot guarantee your complete confidentiality.

A further problem for the idea of anonymity is that the anonymization and de-identification techniques used to protect identities in these datasets are susceptible to being breached. For example existing accepted best practices for de-identification of genetic and health data have been found to be breakable by numerous studies leading to the possibility (Gallagher, Dube & McLachlan, 2018; Gürsoy et al, 2020). This has led to a claim that 'anonymization of data' is a misleading and unhelpful idea and one which may become increasingly irrelevant (Ballantyne, 2019). And, as data analysis techniques develop, it seems likely that any technique which is privacy safe today will be superseded by technical capacity to identify individuals tomorrow. There are multiple suggested responses to this including turning the focus from 'anonymization of data' to instead accepting that data may never be anonymous but nonetheless making it illegal to identify a person in their health data (Gallagher, Dube & McLachlan, 2018). Others have debated whether only a particular set of certified researchers should have access to such data and whether such limitations on access would unreasonably restrict new developments in diagnosis and treatment.

Second, 'big data' analysis techniques raise questions about what kind of data can be considered to be personal health data. Everyday people share large quantities of smartphone data with companies like Apple and Samsung. Movement data on smartphone can be used to predict health status (Kelly, Curran & Caulfield, 2017). Ware et al. have showed that data passively collected from a smartphone (i.e. without the person having to actively enter or send any data) and via university wifi could be used to reasonably accurately predict behavioural and cognitive symptoms of depression (2020). In this context, it is questionable as to whether the distinction between health and non-health data is sustainable.

Third, the question as to what is 'personal' data is increasingly blurred. If I consent to the storage and data analysis of my genetic data it is not only my genetic data that is stored but, by extension, the genetic data of my family and of other people who come from the same ethnic background. In 2018 in the US, police arrested former police officer James D'Angelo for 13 murders and 51 rapes having traced him through a DNA profile. The profile was matched with profiles in a genetic family tree company called GEDmatch, which allowed the investigators to identify people potentially related to the criminal. This allowed a small list of suspects to be identified and subsequently to the arrest of D'Angelo. D'Angelo himself had not shared his genetic data or allowed it to be stored or processed, but, because family members had done so, he was identifiable. The question as to the use of genetic data for surveillance purposes is one that concerns researchers in a context I which multiple countries – including China and the United States – are developing and using processes for large scale surveillance using genetic and other biomarkers (Moreau, 2019).

To summarise:

- The idea of ensuring data privacy via data minimisation makes it challenging if not impossible to conduct the kinds of research that big data techniques allow; this raises issues about the social good of the research vs the privacy rights of the individual participant
- The idea of privacy was often operationalised via the concept of anonymity. It is questionable if data can be anonymized since
    - Large data sets with multiple variables allow people to be identified by crossing multiple categories

    - Even when the dataset doesn't allow people to be identified in themselves, many people now lead very public lives via social media and so crossing datasets with other public information may allow people to be identified
    - Even if datasets are anonymized to the standards of current data analysis practices, new techniques are always emerging which means an anonymized dataset today may not be anonymous tomorrow
- The idea of privacy is based on the idea that there is something special about health data which requires that it is treated to other kinds of data, but big data techniques now mean that the boundaries between health and non-health data are blurred; so the idea that health data is treated in some special way is questionable
- The idea of consent is that I agree to sharing my own data, but in an age of genetic research, any family member who consents actually shares some of my genetic data without my agreement.

The principle of informed consent to control privacy of health data may, therefore, not be practically realisable. While the principle may have made sense in a world before genetic data and 'big data', in which data principally came from individuals who volunteered to participate in a study for a particular purpose, it may no longer make sense in a context in which more and more research is based on large datasets of pre-existing data. In addition, it may no longer be technically possible to readily distinguish health from non-health data, nor to anonymize datasets in a way that will ensure they will remain anonymous.

Should autonomy be the king of principles?

In chapter 7, we noted that the four principles of bioethics were codified in a specific context: that is, they were driven by Western democratic and capitalist countries, in a context in which

individual liberty was prioritized and in which bioethics were seen as a form of 'consumer rights' in which a patient (customer) would be free to make their own choices in the market if they were empowered by perfect information. Costello (2003) has argued that the focus on informed consent as the centrepiece of ethics practices reflects the way in which individualism has become more and more dominant as an ideology within Western liberalism more generally (D'Agostino 1998, Rose 1999). This has led some to argue that autonomy, and the practices of consent, are now effectively treated as the king of bioethics principles:

> …issues of consent get disproportionate air time in the research ethics literature; and even more so in practice where research ethics committees (RECs) or institutional review boards (IRBs) focus primarily on consent forms and participant information sheets (Ballantyne, 2019: 358).

Even if consent is still regarded as central to asking people to agree to participate in research studies which involve subjecting them to particular treatments, it maybe that the traditional approach to consent to use data is questionable in a context of (a) big data research techniques potentially changing the balance between social benefit and individual privacy rights and (b) the concept of privacy itself being increasingly problematic in a data rich world.

Ballantyne argues that the four traditional bioethical principles, which are based in an individualistic perspective should not be the basis on which questions of health data should be judged. She proposes instead the question should be viewed as one of public health rather than one of consumer (patient) protection. In that context, she argues for a different set of principles to be applied to the evaluation of research projects that involve working with biological or genetic data. Rather than being based on protecting individuals, these would be based on shared, social oversight and accountability of such projects. The principles she propose are:

- Public benefit (is there scientific interest and social value in the data)

- Proportionality (can infringement on privacy be reduced such that researchers do only what is necessary to answer socially important questions)
- Equity and solidarity (are risks and benefits shared across community, and can participation be shown to be an investment in the wellbeing of the community as a whole)
- Trust (are steps taken to ensure that the wider community understands what is being proposed and why)
- Accountability (are there processes to ensure that research actions are justified to the public, that risks, data breaches and errors are transparently disclosed).

Conclusion

We noted at the outset that the principle of autonomy is often framed in terms of *informed consent*. Other key concepts which arise in this context are *privacy* and *anonymity*. It has been suggested that autonomy – and by extension, informed consent – have become the 'king' of bioethical principles in western countries. It has been argued that, in practice, research ethics committees rarely focus on question of justice or non-maleficence but rather are principally concerned with informed consent.

In practice, there is evidence that 'informed consent' is not understood by research participants in the same way as it is understood in legal and philosophical definitions. Rather than being a moment of rational and informed decision-making, it appears that the social context of informed consent impacts on the process in a number of ways:

- People are often not terribly well informed by the process about risks or about research procedures as their prior assumptions shape what they learn from it
- People may not think of informed consent as being a decision-making process

- The legalistic form (signing a formal document) may make people feel contractually bound and so actually make people less likely to assert their rights.

This suggests that if we really want people to give informed consent, we may need to adopt a different approach which is based on understanding the participant's perspectives and prior beliefs (caring about), engaging with them over time (caring for), and responsiveness to the development of their understanding of the research process (care receiving).

One of the things people are asked for consent for is the collection, storage, treatment and reporting of their data. This concept is increasingly complicated in a world of 'big data'.

- 'Big data' processing enables new kinds of health care (personalised health care, new diagnostic methods and treatments etc.), and thus may change the balance of 'social good' to 'individual privacy'
- At the same time, the risks of the use of such data for surveillance (by the state or by large companies) raises questions about the balance of beneficence and non-maleficence
- Concepts like 'anonymization' of genetic data which perhaps made sense in the period form 1960 to 2010 may no longer make sense
- Concepts like *individual* consent for storage of data may no longer make sense

In this context, some researchers, like Ballantyne (2019) have argued that we need to move towards practices which are based on public and democratic trust and solidarity, rather than on individual consumer/patient protection. At present, however, Swiss law is framed in terms of concepts developed in a period before 'big data' techniques developed.

References

Ballantyne, A. (2019). Adjusting the focus: a public health ethics approach to data research. *Bioethics*, *33*(3), 357-366.

Beauchamp, T., L and Childress, J.F. (1989) *Principles of Biomedical Ethics*. Belmont: Wadsworth Publishing Company.

Bergler, J., Pennington, A., Metcalfe, M. *et al.* (1980) Informed consent: how much does the patient understand? *Clinical Pharmacology and Therapeutics*, 27, 4, 435–40.

CBS News (2023) Family of Henrietta Lacks settles HeLa cell lawsuit with biotech giant, lawyer says, *CBS News* (1 August 2023) https://www.cbsnews.com/news/family-of-henrietta-lacks-settles-hela-cell-lawsuit-thermo-fisher-scientific/

Corrigan, O. (2003). Empty ethics: the problem with informed consent. *Sociology of health & illness*, 25(7), 768-792.

D'Agostino, F. (1998) Two conceptions of autonomy, *Economy and Society*, 27, 1, 28–49.

Devlin, H. (2025) At-home saliva test for prostate cancer better than blood test, study suggests. *The Guardian (*9 April 2025). https://www.theguardian.com/society/2025/apr/09/at-home-saliva-test-for-prostate-cancer-better-than-blood-test-study-suggests

Estey, A., Wilkin, G., Dossetor, J. (1994) Are research subjects able to retain the information they are given during the consent process? *Health Law Review*, 3, 37–41

Gallagher, T., Dube, K., & McLachlan, S. (2018). Ethical Issues in Secondary Use of Personal Health Information. *IEEE Technology Policy and Ethics*, *3*(3), 1-5.

Gürsoy, G., Harmanci, A., Tang, H., Ayday, E., & Brenner, S. E. (2018). When biology gets personal: hidden challenges of privacy and ethics in biological big data. In *BIOCOMPUTING 2019: Proceedings of the Pacific Symposium* (pp. 386-390).

Kelly, D., Curran, K., & Caulfield, B. (2017). Automatic prediction of health status using smartphone-derived behavior profiles. *IEEE journal of biomedical and health informatics*, *21*(6), 1750-1760.

Lewis, J., & Graham, J. (2007). Research participants' views on ethics in social research: issues for research ethics committees. *Research Ethics*, *3*(3), 73-79.

Masters, J. R. (2002). HeLa cells 50 years on: the good, the bad and the ugly. *Nature Reviews Cancer*, *2*(4), 315-319.

Moreau, Y. (2019). Crack down on genomic surveillance. *Nature*, *576*(7785), 36-38.

Nature (2020). Science must right a historical wrong. *Nature*, *585*(7), 10-1038.

O'Connell, P. (2016). *The battle*. Penguin UK.

Rose, N. (1999) *Powers of Freedom*. Cambridge: Cambridge University Press.

Saunders, B., Kitzinger, J., & Kitzinger, C. (2015). Participant anonymity in the internet age: from theory to practice. *Qualitative research in psychology*, *12*(2), 125-137.

Slevin, M., Mossman, J., Bowling, A. *et al*. (1995) Volunteers or victims: patients views of randomised cancer trials, *British Journal of Cancer*, 71, 1270–74.

Swan, D. J., & Collins, M. A. (2008). Sign on the Dotted Line: The Informed Consent Process (ICP) as Induced Compliance 1. *Journal of Applied Social Psychology*, *38*(11), 2637-2647.

Ware, S., Yue, C., Morillo, R., Lu, J., Shang, C., Bi, J., ... & Wang, B. (2020). Predicting depressive symptoms using smartphone data. *Smart Health*, *15*, 100093.