Good science at heart

Lessons on ethics from big data research in sudden cardiac arrest

Bak, M.A.R.

Publication date
2021

12 | General Discussion

"Begin at the beginning," the King said, very gravely,
"and go on till you come to the end: then stop."
(Lewis Carroll, *Alice's Adventures in Wonderland.* 1865; p. 98)

The final chapter of this thesis reflects on the main findings arising from the research presented. What it means to conduct 'good' science with large-scale datasets obtained from people suffering a cardiac arrest, is discussed first. I then reflect on the research methodology and lastly, provide recommendations for practice and for further study.

12.1 Reflection on main findings

This study addressed the ethical aspects of data-driven research in the field of sudden cardiac arrest (SCA), a condition which is deadly within minutes if left untreated, thereby causing approximately 20% of natural deaths in Europe.[1] SCA is a complex condition, characterized by a multitude of causal factors and by great variation in resuscitation policies between regions. Informed strategies to improve SCA prevention and treatment require the creation of large-scale research databases, which is done in Europe through the ESCAPE-NET consortium formed in 2017.[2] However, collection of clinical and other personal data about this vulnerable patient group, of whom the majority does not survive, exacerbates ethical questions around issues such as data privacy and informed consent. Much had been written about the ethics of *experimental* research in SCA,[3,4] but there existed a lack of literature and practical guidance regarding responsible *observational* research in the medical emergency setting.

Therefore, the overall aim of this research project was to gain understanding of the ethics of (big) data research in SCA and to provide empirically informed suggestions for responsible research with data from people struck by SCA or similarly acute and critical conditions. As described in Chapter 1, the framework of empirical ethics was used in the research presented in this thesis. Our case study consisted of the research registries and genetic biobanks partaking in ESCAPE-NET. A summary of findings from the various empirical and theoretical studies can be found on p. 230. In this section I reflect on those findings and relate them to broader scholarly discussions about big data and research ethics to describe what *good science* means here. This is done in three parts concerning: the nature of these data (12.1.1), the proper ethical and legal governance (12.1.2) and the responsible translation into care (12.1.3).

12.1.1 Good data

Scientific research is impossible without the collection of data. In order to produce knowledge from these data that may be used to benefit future patients, it is vital that the data are of good quality and that a suitable study design is used. The use of big data, defined in Chapter 1 according to the different v's, amplifies the limitations of traditional data science.[5] In the emergency research setting quality control has always been very important as it may be chaotic, and because there is great variation in registration and reporting of SCA which complicates harmonization of databases, despite the existence of the Utstein standard.[6-8] These issues are merely intensified with big data, so the findings in this thesis are equally valuable for 'small' data research.

The use of resources and the risks to people's privacy can only be justified when there is a valid methodology. Incorrect usage and missing or low quality data can have real-world impact for patients. Observational studies are prone to confounding and other biases, however. One example is 'resuscitation time bias' where patients receiving interventions that are usually implemented relatively late in the resuscitation effort (such as endotracheal intubation or administration of epinephrine), are more likely to have a longer cardiac arrest: these interventions will therefore be associated with worse outcomes, if not corrected for or standardised for the duration of the resuscitation effort.[9] Also, in large-scale SCA research like ESCAPE-NET, there is a risk of selection bias due to differences in policy contexts and regulations. When data of deceased patients may not be used, the results are skewed towards surviving individuals; while requiring survivors' consent can artificially lower survival rates when data from non-consenters are removed. The collection of at least a minimal dataset from all patients is thus necessary for drawing valid epidemiological conclusions.

Moreover, researchers should be aware that correlations in data are valuable only against the background of a certain theory, or in other words, a certain context. Luckily, there is increasing recognition among philosophers and data scientists alike that 'raw data' is an oxymoron.[10] Data are theory-laden, cooked rather than raw: we know since Kuhn that data are not 'out there' to be found, but that they are constructed by people within a particular socio-historical context. When existing *societal biases* are not corrected for, these can be reproduced through big data analytics, e.g. machine learning methods that replicate gender or race stereotypes.[11] The choice of in/exclusion of data from certain communities is an ethical issue as well. Results from European projects cannot be applied directly to regions with other ethnic populations or differently structured emergency medical services, so local research is needed across the globe and with diverse groups of participants.[12] This was seen in discussions about a Dutch guideline for decision-making about resuscitation, in which worrying international statistics about neurologically favourable SCA survival among the elderly were presented which are not representative for the Netherlands and might scare seniors

into do-not-resuscitate orders.[i] Another example, discussed in Chapter 2, is the lack of inclusion of African-Americans in research which led to misclassification of common genetic variants so that this group was wrongly thought to be predisposed for hypertrophic cardiomyopathy.[13] Such mistakes can intensify existing health disparities and increase stigmatisation of minorities.[14] Conversely, data obtained from underprivileged communities should be used to benefit these same communities rather than benefiting only affluent groups. The citation opening that chapter was about Henrietta Lacks, a black woman who died of cervical cancer and whose tumour cells ('HeLa cells') were used and sold by researchers without the knowledge or consent of her surviving children, who struggled to afford health insurance. Her story shows that researchers need to keep in mind that research data (and tissue) stem from actual persons, rather than seeing these as two divorced entities. Systematic collection of personal data may lead to researchers adopting the "objective" stance, in other words, a detached attitude to the data subject.[15] Health datasets are growing bigger and bigger, but we should not forget the individuals constituting them. What this means for the governance of health data research, I discuss later in this chapter.

Not only the selection of persons about whom data are collected, but also the type of data that is collected, has moral relevance. We saw in Chapter 3 that when data are collected outside of the medical context, such as socio-economic data, patients desired more knowledge of and control over the uses of these data. This makes sense when viewed from Nissenbaum's theory of privacy as contextual integrity: privacy always functions within a specific context, or Walzerian 'sphere'.[16,17] When data move between spheres, data subjects may become concerned and hesitant to share. This is why we argued in Chapter 6 that it does not make sense to isolate specific types of data as exceptionally sensitive. Instead, their sensitivity depends on the particular context. Genetic data can be both more and less sensitive than HIV test results, for example. We advocated the framing of *contextual exceptionalism* as alternative to the stubborn notion of genetic exceptionalism. Contextualism provides a more sustainable framework for the quickly changing world of health data research; and when combined with universal human rights norms that are widely held across societies, contextualism does not amount to moral relativism.[17]

Embedding context-sensitivity in health data research requires recognition of the messy nature of reality which may not be easily captured in a medical dataset. Some people still see biomedical data science and social science as strictly separated domains, however. In a recent webinar, humanities scholars who authored a book called 'Your computer is on fire' shared an insightful anecdote. They had invited a university computer science department to a conference about technology and society, but the

[i] Alwin Kuiken, 30 juli 2015. Trouw. 'Richtlijn voor reanimatie negeert situatie Nederland' [Resuscitation guideline ignores Dutch setting]. [https://www.trouw.nl/nieuws/richtlijn-voor-reanimatie-negeert-situatie-nederland~b2cec5c2/]

department chair had said he was only interested in how computer science can improve society, not in how societal aspects should be considered in the development of computational systems. Such technological utopianism glorifies the possibilities of big data analytics and disregards the importance of incorporating aspects that might escape *datafication*. This is perhaps inherent to big data research, which is typically grounded in the paradigm of digital ontology. According to that view reality is ultimately computational, made of 1s and 0s. However, rather than digital, a broader ontology grounded in information is more useful. Namely, information can be presented in digital and non-digital form (the Second World War population records, mentioned in Chapter 1, had not been digitalised), and information can be binary or non-binary (suffering cannot be expressed in 1s and 0s and arguably not as continuous variable either, not without losing context). Digital and analogue are merely different glasses, frames, or levels of abstraction for viewing the world.[18]

Outside the frame, is what we're leaving out.[ii] Big data researchers risk missing out on important qualitative information if their frame is one of numbers. Only quantifiable experience matters: to be, is to be computed.[19] One example is the quantified-self movement where people use quantifications to make their bodies more predictable.[20] At population level, quality adjusted life years (QALYs), an economic measure for disease burden, have been criticised because the value of a person's life should not depend on contingent factors such as healthcare budgets, and because the quantitative nature of QALYs does not account for contextual factors such as baseline health status or the patient's values and preferences.[21] By prescribing which treatments are reimbursed, QALYs reduce autonomous choice. Datafication, when understood as quantification, may thus lead to a false sense of certainty and objectivity while having real-world consequences. When health interventions whose effectiveness cannot be quantified through traditional methods are not covered, important values in care may be lost.[22] Instead, health data researchers should combine qualitative data with qualitative research methods "which do justice to the complex outside world" (p. 21).[23] Namely, when patients are reduced to what makes them different from others (healthy = 1, unhealthy =0) this limits their possibilities and opportunities for change:

> "The world breathes, its heart beats, like the heart of computer technology beats between 0 and 1. Does the breathing cause the beating or does the beating propel the breath? In the beating of the heart, systole and diastole alternate, the contraction and relaxation of the heart chambers. There is no presence and absence, but continuous movement (…), and to humans this movement quite literally means *the possible* as Derrida describes it – the movement of the maybe, the *peut-être*." (p. 94, translated from Dutch)[24]

[ii] Queens of the Stone Age. *Go With The Flow* [Songs for the Deaf]. Interscope Records; 2002.

12.1.2 Good governance

Data-driven research can lead to more personalized care, potentially while bringing down costs, but it may also require sacrifices. The studies in this thesis found that SCA data research brings instrumental and principled risks, mainly to privacy, but stakeholders agreed that the risk of harm to subjects is low compared to intervention research. I did not aim to enumerate the harms to participants in data-driven projects like ESCAPE-NET, but to provide an overview of areas of ethical concern. Further study would be valuable that maps the actual extent of harm due to big data research in medicine. Interviewees saw few risks, but researchers and oversight bodies should remain vigilant as the literature suggests that the harms of big data are difficult to anticipate and that cyber-attacks will become increasingly prevalent in health organisations.[25-27] The growing technological possibilities associated with big data should thus be accompanied with a growth in knowledge on governing data-driven research.

In Chapter 4, we saw how privacy by design ('hard measures') and privacy by policy ('soft measures') are needed to protect patient data. Figure 12-1 shows the current paradigm for data protection in observational studies, where less sensitive and more anonymous data come with fewer governance requirements. The GDPR does not cover anonymous data at all, and has special provisions for health data. When sensitive personal data are used for research, privacy by design measures such as data minimisation or privacy-enhancing technologies (PETs) are supplemented with policy measures like informed consent procedures. However, my findings suggest this framework is outdated. Namely, interviewed experts and literature confirmed that complete de-identification is impossible, especially in SCA research where the chain of care results in many different data sources. Complete anonymization might also be undesirable because it can affect data quality, limit options for audit trail, prevent the addition of new data to a profile and remove the possibility of contacting the data subject (including giving people access to their personal data).

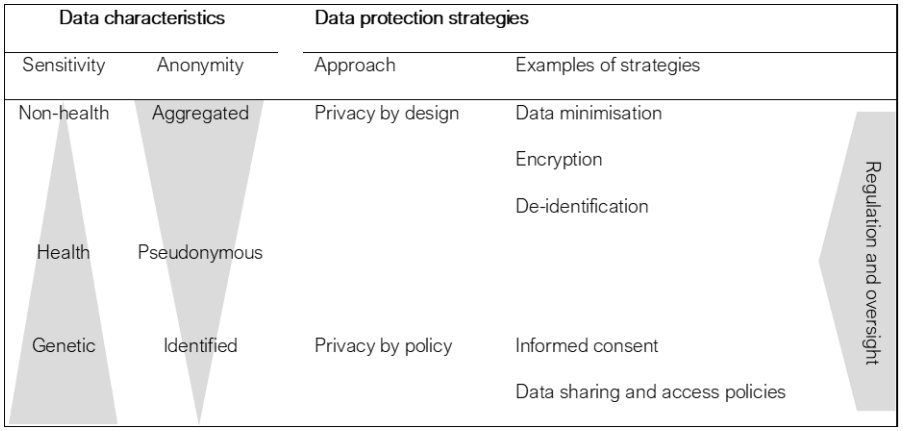

Figure 12-1. Current governance framework for health data research. *Figure based on Hoepman, 2014 and Spiekermann & Cranor, 2009.*[28,29]

Regarding the sensitivity of data, I discussed in Chapter 6 and in the previous section that genetic data are not more sensitive *per se* but that sensitivity depends on context.[iii] Deep phenotyping studies also create a highly detailed individual profile, and even data collected outside the medical context may be used to infer information about someone's health: for instance, web users might be identified by typing patterns after which their speed and accuracy serve as a predictors for psychological distress.[30,31] This is more acceptable if done for health research purposes e.g. in order to create an empathic psychotherapy chatbot, than when used for targeted advertising. My point is that less anonymous and more sensitive data indeed require stronger protections but that, unlike what Figure 12-1 suggests, sensitivity and anonymity cannot be pre-determined. The characteristics of research data, which also include data validity as discussed under the heading of 'good data', vary according to the specific context.

Context also affects the values and norms operating in the background of governance decisions. If underlying values are not explicitly referred to, the normative background may be forgotten which can lead to routinisation of researchers' data protection practices and cause resistance to regulations thought to harm the research enterprise (Chapter 11). In my study of ESCAPE-NET, researchers dreaded the lack of clear and uniform guidance. Their experience may be reflected on using the Lewis Carroll citation at the opening of this general discussion, taken from a chapter titled *Alice's Evidence* which details the criminal trial of the Knave of Hearts. He is accused of stealing the King's tarts, but the evidence presented at the trial turns out to be a nonsensical poem supposedly written by the prisoner. The King demands that this poem is read from beginning to end, while failing to explain its meaning. Still, the poem is considered important evidence and the Queen calls for sentencing of the Knave – even before the verdict arrives. This random turn of events goes against all Alice's intuitions about justice in her own world. Carroll contrasts the arbitrary logic of Wonderland with the ordered legal system in nineteenth-century Britain, where the reverence of scientific principles reflected a desire for certainty. In a manner similar to the legal process in Wonderland, albeit much less extreme, data protection laws such as the GDPR are necessarily open to interpretation, as this is needed to allow for changing contexts and differing societal values.[32] The different interpretations, however, frustrated ESCAPE-NET researchers who desired certainty.

In response to the strict (interpretation of) data protection rules some scientists, policy-makers and bioethicists, have argued for the recognition of solidarity as the foundation for health data research, as described in Chapter 11. Yet the studies in this thesis have shown that neither solidarity not privacy can function as foundational value. The search for final principles leads to an infinite regress, as Winch states when

[iii] Two factors that can make other health-related data as sensitive as genetic data are the relationality ('do the data also create privacy concerns for other people?') and the embodiment of the data ('are the data constitutive of the person's physical and/or informational identity?').

he reflects on another work by Carroll ('What the Tortoise Said to Achilles'): "although we may formulate another, higher-order, set of precepts prescribing how the first set is to be applied, we cannot go further along this road without finding ourselves on the slippery slope pointed out Lewis Carroll" (p. 55).[33] We showed that the fundament of rules related to health data research is suspended between societal actors who are bound together through *trust*. We argued that instead of imposing stricter or more rules, trust should be explicitly recognised as the background condition for governance.[iv] My proposal for a framework that recognises governance as context- and value-dependent and based on trust, and that replaces the one in Figure 12-1, is depicted in Figure 12-2. As depicted by the arrows, promoting trust requires that contextual integrity is respected (i.e., expectations of privacy in a certain context are met), or more specifically, through the use of governance measures that are appropriate to the particular characteristics of the data. When the link between context and values/norms is severed, trust disappears as basis for health data research.

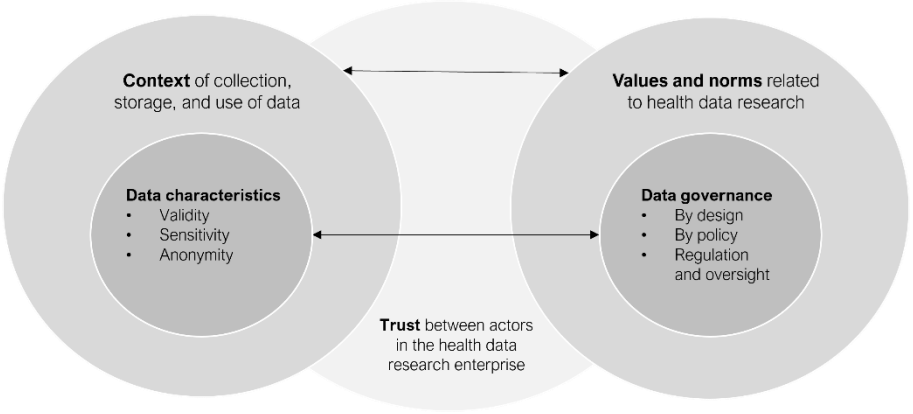

Figure 12-2. Proposal for a new governance framework for health data research

Thus, when deciding on the data governance measures needed, these links and the effects on trust should be taken into account. This may be illustrated with the ongoing discussions on the need for consent in health data research. Informed consent is a key principle in research ethics and relevant law but requiring consent may lead to non-use of data, which conflicts with the duty of beneficence and with individual autonomy if it violates people's wishes to contribute to research. Some scholars think that this creates justification for a general waiving of consent for health data research.

[iv] I think the terminology of *data governance* is more useful than the narrow term *data protection*, because it also includes principles like accuracy and accountability which are not directly related to securing who has access to the data. This interpretation is in line with the European Commission's legislative proposal for a Data Governance Act to facilitate data sharing.

Among others, Porsdam Mann and co-authors advocate a duty to participate in 'minimal risk' health data research, by arguing that the burden and risk of (instrumental) harm is lower than for other societal duties, such as the duty to quarantine in case of dangerous communicable diseases.[34] I disagree with this conclusion. The link between data donation and benefits to others is far from direct, while privacy is important not only in instrumental but also in principled terms. Perhaps people have an imperfect moral responsibility to participate in projects like ESCAPE-NET, but whether they do so is *their* business and should not discharge researchers from engaging in good governance: the relation towards the data subject is a non-symmetrical relation.[35] Even for widely supported research, minority groups can have objections based on religion and individuals may have private reasons to oppose a certain practice (e.g. bad experience with the hospital where the research is conducted). Existing public trust is thus not a reason for removing the possibility for opt-out. Moreover, mandating data donation would itself negatively impact on trust. I contend that the respect for data subjects' autonomy and the promotion of trust, together create an important *prima facie* rule for requiring ('opt-in') informed consent from patients whenever possible, such as in the ARREST study described in Chapter 3.

This point remains true even if SCA survivors could not reproduce their consent, which is in line with Nissenbaum's transparency paradox: while autonomy is best preserved by long and detailed information, no one will read, understand or remember it.[16] Yet consent of these patients is still valuable, as the value lies in the fact that they can refuse when they do not trust the researchers. Nissenbaum explains that in health care and research where trust is generally high "it is not the consent form itself that draws our signature and consigns us to the operating table, but rather our faith in the system" (p. 36). This system is characterised by a governance framework of "supporting assurances". Health data research should ideally employ a combination of privacy by design and privacy by policy measures, together with ethical and legal oversight such as by RECs and DPOs. Namely, ethics oversight is needed so that the burden of privacy protection does not lie entirely with patients (whose options for control are limited) and with researchers (who have an inherent conflict of interest)[38]. Moreover, approval of the study by an REC serves as a 'trust-tag' for scientific partners and data sources, and thus contributes to successful data sharing.[37]

Regarding deceased data subjects, the current lack of regulation and guidelines about post-mortem research uses of data may start to harm public trust as personal data are accumulating online, e.g. through memorialised Facebook pages or cause of death registries. As opposed to data, tissue has always been seen as constitutive of identity. In a documentary about HeLa cells, one of the family members says: "Henrietta ís those cells". Now the data shadows of the dead are becoming increasingly sophisticated informational entities – yet ones who cannot control the future unexpected uses of their data. Floridi describes how the distinction between online ('here')

and offline ('there') is blurring fast.[38] This reminds of the futuristic television series Black Mirror that featured an episode called 'Be Right Back' where a grieving woman buys a synthetic AI-based recreation of her dead boyfriend. If there is some truth in this episode, then in the near to long-term future deceased persons may be *here, there and everywhere*. The studies in Part III of this thesis conclude that a moral right to post-mortem privacy should be recognised and acted on using the same framework that applies to the living, although their interests may outweigh those of the dead in certain situations. In studies for which consent has been given during a person's lifetime, it would be acceptable to assume that this consent remains valid after death, but ideally the post-mortem use is specified in consent documentation. In the same way, careful consideration of the interests of the deceased and their relatives is needed before identifiable case reports can be published (Chapter 7). Of note is that while case reports have been used for scientific research and education since antiquity, the trend towards incorporating extensive data about social determinants of health may require future changes to the privacy standards described in that chapter.[39]

Lastly, in situations where implementing certain governance measures would limit the conduct of research, competing values need to be balanced. Following Resnik, I stated in Chapter 11 that trust should be used for priority-setting between values of research ethics, but that it is not a "meta-rule" that always trumps other principles.[40] Only in a world where everyone has perfect understanding of context, would trust be justified as meta-rule. For instance, even if proper safeguards are in place, people may not trust research using commercial data processors (e.g. companies that store data or provide computing power) due to limited grasp of the role of these parties. Deciding on the consent process then requires balancing trust with beneficence. In another situation, people's trust in researchers may be misplaced and conflict with non-maleficence.[v] Resnik therefore includes the proviso that people's trust should be based on reasonable expectations of good conduct. In both examples, using the principle of trust brings out a requirement for good communication and discussion about these principles between potential participants and researchers. Moreover, we already saw that trust may be different for individuals than for society as a whole.

12.1.3 Good care for all

The title of this thesis demands a teleological explanation, as we need to ask: good science for what/whom? Research data are not good or bad as such. Their goodness depends on how they are collected and governed, but moreover, on how they are *used* in healthcare and beyond. Who benefits from this use? Hereafter I discuss how data-driven SCA research is (in)directly linked to care at individual and societal level.

[v] Resnik's example is the 'Tuskegee Study' that withheld treatment from participants.

Participants' trust in researchers was based on their trust in the medicine. The boundary between research and care has become blurred, which is observed especially in discussions on the return of genetics findings. Chapter 3 provided evidence of a 'therapeutic non-conception'[41] in that subjects were unaware of genetic research and the corresponding potential of receiving individual findings – in other cases, their preferences changed. Decisions not to know are not always well informed and the *right not to know* may be trumped by other interests.[42,43] Researchers, who are often also clinicians, may experience moral distress when the option to potentially save a life is removed (Chapter 4). This thesis did not focus specifically on genetic studies and does not provide conclusive results about the disclosure of individual genetic findings, so I can only make the tentative suggestion that for life-threatening but potentially preventable conditions such as SCA, a general right not to know is too demanding on researchers if data subjects are identifiable.[vi]

My suggestion is in line with the policy of our own hospital which prevents patients from participating in genetic studies when they do not want to be informed of individual findings. However, rather than instating such a general policy, I suggest that oversight bodies should evaluate the contextual factors for each proposed study (e.g. the clinical significance of finding or seriousness of the harm) in a manner analogous to the framework presented in Figure 12-2. This suggestion is underwritten by empirical survey data where 96% of IRB members supported the right not to know, which dropped to only 35% when a specific scenario was presented in which disclosure results in prevention of serious disease.[42] Ethics committees should similarly discuss the return of findings to relatives of deceased research participants. Further study and guidelines on post-mortem disclosure are needed, as was concluded by the European Society of Human Genetics in their reflection on Chapter 5's findings.[44]

Once it is determined that research findings should be shared, these need to be made available to a healthcare professional for disclosure to the patient or relatives, which requires good communication between the research and care settings. This also applies to AED data which only become available to the treating physicians because they are collected by researchers, even though these data may be important for diagnosis and treatment (Chapter 8). The data can also help first-responders to improve their CPR performance but are not collected routinely. (In some European countries these data cannot become available at all because AED use is reserved by law for medical professionals, e.g. Italy.) Standardised collection of AED data for healthcare would remove the reliance on researchers. Policy debates are needed about which information related the SCA event should be included in the electronic health record, such as the AED data, and on the needed infrastructure.

[vi] My point applies to findings related to the studied condition (SCA), the return of entirely incidental findings is beyond the scope of the current discussion.

Moreover, the policy and regulatory debate on health data research focuses mainly on individual safeguards such as consent and anonymization, while big data enlarges the potential for group-level harms.[25,45] Disparities in SCA care have always existed, e.g. people in rural areas are less likely to benefit from quick EMS response than those in urban areas,[46] and women are less likely to be resuscitated than men.[47] However, the technological potential of data-driven research makes humanity more responsible for the outcomes. The optimistic explanation of this is that digital tools help bring injustices to light: studies about women's lower chances of being resuscitated have prompted media attention which creates societal interest in (funding) further research on this topic; and the visualisation of AED distribution across regions might help to show inequities and fuel action aimed at equal access. With this goal in mind, the Dutch Heart Foundation published an online tool where people can check whether an AED is still needed in their area (Figure 12-3).[vii] Another new initiative is drone-based delivery of AEDs.[48] Complete coverage would be ideal but is unlikely to be realistic under budget constraints, given that these devices are expensive. Public health initiatives aimed at improving the use of AEDs therefore increasingly utilise computer models to simulate what the optimal distribution would be, as described in Chapter 10. However, this requires programmers to weigh equity with efficiency; i.e. saving the most lives versus providing an equitable chance of being saved.

Figure 12-3. Screenshots from the Dutch Heart Foundation's website showing: (a) the locations in the city of Amsterdam where no AEDs are registered with the citizen responder dispatching system 'HartslagNu' that are available day and night; (b) the search result which shows that a 24/7 AED is still needed in my own neighbourhood.

[vii] Dutch Heart Foundation (accessed 22 June 2021). Mist jouw buurt nog een AED? Check waar AED's nodig zijn. [Is your neighbourhood missing an AED? Check where AEDs are needed]. [https://www.hartstichting.nl/aed]

This bring us to the more pessimistic interpretation of justice aspect of data-driven research. In addition to the arbitrary decision-making regarding equity, described in Chapter 9, models may be unjustly biased against groups as a result of distortions in the training data. A much used example is *predictive policing*, where racial bias arises as algorithms encourage police patrols to target certain ethnic groups.[49] In the Netherlands, big data collection led to citizens receiving childcare benefits being wrongly labels as fraudsters, thus pushing families into financial hardship as a result of failing oversight and limited transparency.[viii] A third example focuses on language use. When translating from gender neutral languages that only have one word for he/she, the algorithm in Google Translate systematically changes the gender of translations to fit with stereotypes: 'he is a doctor' and 'she is a nurse'.[50]

In the same way, prediction models in health care may be unjustly biased. This bias may be propagated if not corrected, especially when big data analytics are used. For instance, when predictions about certain groups of critically ill patients lead to withholding of treatment on the basis of poor prognosis, this may result in a self-fulfilling prophecy.[51] While a critical doctor might recognise false-positives based on contextual factors, computer models are generally not advanced enough to do so, and are not always transparent enough about the factors for prediction to be checked. Mertens & King detail how artificial intelligence (AI) based models for neurological prediction after resuscitation exacerbate this potential for self-fulling prophecies, by creating feedback loops that lead to increasingly negative outcomes among (certain) SCA patients in the intensive care unit.[52] The PROFID project introduced in Chapter 10 aims to create and clinically validate an AI-model to determine ICD implantation but runs the risk of withholding ICD treatment to patients misclassified as low-risk, if the model were to be trained on inaccurate or biased data.[ix]

This theme is again reflected in Lewis Carroll's work: in the 1871 sequel to Alice's Adventures in Wonderland titled 'Through the Looking-Glass', Alice finds herself in a country that is bound by rules, like a chessboard where the behaviour of the pieces is captured in the rules of the game. The red and white queens of this country want to punish people before a crime has been committed. As Siemann notes, this actually

[viii] Harry van de Loo, Ben Dankbaar, Sjoerd Romme. 10 June 2020, Trouw. De toeslagenaffaire bij de belastingdienst is het actuele voorbeeld van een oud probleem. [The childcare benefits scandal is the contemporary example of an old problem]. [https://www.trouw.nl/opinie/de-toeslagenaffaire-bij-de-belastingdienst-is-het-actuele-voorbeeld-van-een-oud-probleem ~baa80e1f/]

[ix] If we take a global health perspective, the model might not be valuable for countries where certain predictors are not available. Namely, among PROFID researchers discussions are ongoing about whether magnetic resonance imaging (MRI) data should be included in the prediction model. Findings from MRI scans are associated with SCA incidence and function as important predictor in the model that determines who receives an ICD. However, the consortium also includes a number of partners from countries where hospitals do not generally have MRI machines, as these are quite expensive.

makes sense in the chessboard world where action is bound within the game's limits.[53] The queens are like AI-based prediction tools such as the ones described above: within the parameters of the model, and based on inferences where data are missing, the data subject is classified as high-risk and treated as such.[x]

While risk prediction is beneficial at a societal level and potentially also for the individual, we should not forget that when patients are reduced to data entries, this can harm the respect for individual freedom and human dignity. SCA patients may consent to the use of risk prediction tools because they fear consequences for their care, that is, for their life – just like Alice consents to be governed by the red and white queens as she has a vested interest in cooperating with them.[53] Consent is never entirely free as it is shaped by social forces and, as Foucault described, biomedicine can create an abusive *power* structure.[54,55] From a phenomenological perspective we can see how the representation of patients into predetermined data categories means "bringing the other under the power of the same", to put it in Levinasian terms.[35] Data-based tools create an aura of certainty, but the acute and critical care setting remains characterised by uncertainty and dilemmas. In reflecting on the challenges around prediction of prognosis in the intensive care, Wilkinson concludes that "what we ought to do in this case or in other situations of uncertainty depends upon the *value* that is placed upon different outcomes, including how bad it is to prolong treatment that ultimately leads to a bad outcome compared to how bad it is to withdraw treatment from a patient who would have survived without impairment"(p.408).[51]

This focus on values leads to another societal issue, briefly mentioned in Chapter 4: should we accept SCA as a good death in certain populations such as the elderly? Sociologist Stefan Timmermans argues that CPR is part of a society that medicalises death and corrupts the dying experience.[56] According to Timmermans, resuscitation research turned death into something that could be overcome, even though most victims do not survive.[xi] He critiques the performing of CPR in medical settings on patients who are unlikely to regain consciousness and writes: "in its pure form, sudden death is very close to our contemporary conception of a good death, while in its current form it's a death to be avoided" (p. 28). Namely, while the in-hospital resuscitation allows next-of-kin to conclude that everything possible has been done and helps them adjust to the death, it might also de-socialise the dying process and harm the dignity of the person receiving CPR. Instead, we should focus on dying with dignity and on living our best lives: as Nietzsche and Heidegger already said, confronting our death anxiety creates the opportunity for finding meaning in life. This

[x] The difference with our world is that Looking-Glass country is ruled by women, while in reality authoritative industrial AI applications are represented as male (IBM Watson) and powerless and submissive personal assistants (Apple's Siri and Amazon's Alexa) as female.

[xi] A parallel can be drawn with *dataism* which also gives the illusion of immortality as people live on in the cloud, as discussed in Chapter 6 and by Visser-Knijff & van der Vaart.[57]

confronting implies that discussions about death are needed, as well as the promotion or creation of advance directives among older populations to ensure that their wishes are upheld.[58] Timmermans also states that a focus on survival rates distracts from the fact that some patients survive with low quality of life. The PROFID study addresses quality of life, as its non-inferiority trial is justified through the prevention of adverse effects of ICD implantation (Chapter 10). However, in this type of study there always remains a margin of clinically acceptable loss of efficacy: personalizing the decision-making around ICD implantation comes with a higher risk of death in the population who do not receive the ICD on the basis of concerns for their quality of life.

The aforementioned issues are pressing in light of resource constraints. Strengthening EMS infrastructure and placing defibrillators is costly. In 1995, the cost of CPR per quality adjusted life year was calculated at one and a half times the median family income in the US.[56] While cost-effectiveness analyses of CPR are difficult to interpret as they depend on how quality of life is measured,[59] resources might be more effectively invested in the *prevention* of SCA by readjusting the focus to determinants like stress, diet and exercise. Yet from a trust-based and human rights perspective, public health considerations should be constrained by a concern for the rights of individuals, including of those without fair opportunity to benefit from (newly instated) prevention policies.[60] Utilitarian philosopher Peter Singer argues that resources should not be spent on those who cannot contribute to society, but when his mother developed Alzheimer's disease, he spent a lot of money on her care and said: "Perhaps it is more difficult than I thought before, because it is different when it's your mother". While Singer takes this to mean that he failed to live up to his own standards,[xii] how I interpret the anecdote is that policies should leave room to recognise the moral importance of the individual person. In this thesis I do not, however, take a normative stand on the broader societal effects of SCA research, but rather propose that these should be taken into account in future ethics parallel research and potentially also by RECs. Just like the anthropologist Alice studies the worlds in which she is thrown, medical ethicists should pay close attention to the societal impact of big data.

12.2 Methodological considerations

For the past years I was part of the European SCA research consortium known as ESCAPE-NET, wherein this work is situated. The scope of this ethics research was not delineated from the start but rather, most studies followed iteratively from others and were inspired by practice. The journey has not been a straight path and I dare not say that it 'began at the beginning', as discussions about data protection in health

[xii] Daniel Gross, 25 April 2021. The New Yorker. Peter Singer Is Committed to Controversial Ideas. [https://www.newyorker.com/culture/the-new-yorker-interview/peter-singer-is-committed-to-controversial-ideas]

research have been going on since the 1980s, if not longer.[61] The end is not in sight either, because all topics in this thesis require further investigation and have practical implications that will not be solved quickly.[xiii] What this thesis does, is use observational SCA research as a critical case from which the ethical and policy implications can be analytically generalised to other types of illness, as described in Chapter 1.

The scope of the thesis allows for providing a comprehensive overview of the diverse aspects associated with observational SCA research, but herein also lies an important limitation. Namely, due to the variety of ethical issues arising in a large project like ESCAPE-NET and the lack of focus on one particular issue, my findings have remained somewhat superficial: topics such as scientific validity, data sharing with commercial partners, or the return of genetic findings, could each fill a doctoral dissertation on their own. The limited specificity of the research questions resulted in large volumes of qualitative data that were sometimes difficult to navigate and of which parts remained unused. However, these data may be valuable for re-analysis on specific topics.[62] Regarding the scope of the research, it is also important to note that the main focus was on SCA data rather than on tissue: if biobanks were mentioned, I mostly discussed the genetic *data* extracted from patients' tissue samples. The ELSI of SCA biobanking studies unrelated to the resulting data, such as questions on the control and ownership of the tissue, remain to be discussed elsewhere. In addition, I was unable to compare specific policies of the different ESCAPE-NET partners due to the high heterogeneity of the studies and the rarity of events within them, e.g. for drawing conclusions about the correct manner of approaching parents of children who died from SCA a much more specific approach would be needed.

Throughout this chapter I referenced Lewis Carroll, whose work resists interpretation in the sense that we cannot assign universal meaning to the events in his books. This PhD thesis is like Alice's adventures, because while I try to make sense of data ethics, it turns out that sensible things can only be said in the context of a specific practice. The articles in this thesis have not promoted a particular normative standpoint that is *a priori* true, but are a collection of adventures that inform normative recommendations *a posteriori* (see 12.3 for such recommendations). As such, this work is grounded in the framework of empirical ethics, which was introduced in Chapter 1. The strength of that approach lies in its sensitivity to context and in the involvement of relevant societal stakeholders.[63,64] Not only the empirical studies but also the conceptual analyses in this thesis (i.e. on the nature of trust and relating to post-mortem privacy) were situated within the specific setting of health data research, as we know since Wittgenstein that the meaning of philosophical concepts is deter-

[xiii] Some findings may have become dated in the course of this project, such as the literature search conducted in 2017 (Chapter 2). The effect on the final thesis, however, was mitigated by keeping up to date with the literature throughout, e.g. with the help of PubMed alerts.

mined by their use in a context.[65] The paradigm employed was a dialogical contextualism instead of the coherentism inherent to the reflective equilibrium (RE) method. This approach goes further than RE in that it recognises practice not just as a source of 'well-considered judgments' but as the setting in which societal actors negotiate about normative claims, and is thus less vulnerable to the charge of conservatism.[66]

For the empirical component of this thesis, I used a combination of literature study and qualitative research methods to explore the views of stakeholders. This gave more insight than a quantitative survey would have, because interviews and observations provided the opportunity to build rapport and to interact with stakeholders, whenever that was useful and appropriate. Semi-structured interviews with SCA survivors were conducted in person, in their homes unless otherwise preferred, which was worthwhile as it enabled obtaining a more detailed picture of that person and their values. This is illustrated by the following quote from one of our interviews:

> Patient: Can I ask you a question? You come here to visit and that gives you a very different impression of the person than when you would ask questions via e-mail or telephone, right? Now you see the seriousness of the situation and can draw conclusions for your research. If you would just have information on paper, that would be different.
> Patient's partner: Yes, if you would do your PhD research on this topic without ever talking to someone who experienced [a sudden cardiac arrest], then your research would be much less valuable than what you have now.
> Me: That is absolutely true. But of course my PhD research is about ethics, and for the interview we were talking about medical research on how to improve cardiac arrest treatment. For those studies you can use statistics, while 'my kind' of research focuses on people and perspectives.
> Patient: Still, they can have the largest possible list of data, but it would be worth much less than this conversation.

This patient and their partner also make an important point about the nature of data-based research in medicine, one that I did not realise fully at the time. Section 12.1 described how the collection of certain data obscures the information enclosed in data that are *not* collected, such as data of qualitative nature. Embedding qualitative research into big data projects does not only shed light on the ethics and acceptability of these studies, but can also improve them scientifically and help to set relevant research priorities.[67] In ethics parallel research, ethicists are part of the projects that they study, as has been the case with ESCAPE-NET. This type of interdisciplinary collaboration has the advantage of allowing co-construction of best practices but may also result in a conflict of interest when ethicists are not able to be critical. It is important for bioethicists to be reflective about their role and to ascertain which

projects are suitable for parallel research and which are not. This also depends on timing, as they should avoid the Collingridge problem of being either too early or too late to significantly influence the development of a technology (such as a database or risk prediction model).[68] In the case of ESCAPE-NET, these limitations were mitigated through transparency about collaboration and financial support – although affiliation with the project under study might still have influenced interview findings – and by involving an independent ethicist in the ethical evaluation of the project. We were also lucky to cooperate with scientists who had an open mind towards ethics. Resuscitation research has always been closely tied to ethical discussions, which may have promoted this collaboration. During the writing of this thesis I also experienced how close the link between ethics and law is, as both consider how proposed norms can be used in practice and whether they are compatible with contextual values, so further collaboration between these fields would be valuable.

Specific study limitations were addressed in the corresponding chapters, but one key limitation arose across several articles: namely, the generalizability of findings. Articles included in literature reviews were few and mostly originated from Western European or English-speaking countries, as did interviewees. This is not problematic *per se* since the ESCAPE-NET project is located in Western Europe and, moreover, because qualitative research does not aim to generalise but rather to provide insight in contextual factors and the conduct and opinions of stakeholders. However, description of these opinions and the identification of otherwise overlooked issues does require inclusion of a *diverse and representative* group of stakeholders.[69] In the surveyed empirical literature and in our own research, there was to some extent a lack of diversity in terms of age, ethnicity, and nationality of respondents. Views of deceased patients' relatives are missing, as well as of patients who declined consent to data research. I also would have liked to include SCA survivors in the roundtable conference, but it was difficult to recruit patient representatives as they worried about the knowledge asymmetry with academic attendees. Further work on this topic could try to mitigate power inequalities by conducting participatory action research,[70] as well as aim for mixed-methods research that also includes quantitative methods in order to enumerate opinions and help weigh priorities.[71]

Lastly, at the time of writing not all papers in this thesis have been published and some are still undergoing peer-review, so these articles might be improved in later versions (Chapters 4, 6 and 11). It may also be noted that a number of articles in this thesis are position papers which could seem divorced from deeper ethical reasoning, but they are actually very much connected to the empirical-ethical work done in other articles. Unfortunately, position papers provide less space to elaborate on the background of recommendations. I hope that in presenting the different types of articles together in this collection, the connections between them have become clear.

12.3 Implications for practice and further study

Goethe already said that *"knowing is not enough; we must apply"*. Hereafter, I suggest how the findings from this thesis, as discussed in section 12.1, can be applied to health data research in practice. Practical recommendations and suggestions for further ethical study are made. It is important to note that incorporating ethics into data-driven research does not necessarily limit the value of these studies, as is often thought. My recommendations are not anti-technology but recognise the trade-offs inherent in research (as in any other social practice). Since Hobbes we know that even having a government can be considered a trade-off: citizens trade a piece of their individual freedom for guaranteed protection from a violent 'state of nature'. In the same way, privacy is a *prima facie* right that should be balanced with the benefits of science. What should be done to promote SCA data research while safeguarding subjects' rights (12.3.1) and how can such research be evaluated in a fair and just manner (12.3.2)?

12.3.1 Empowering effective and equitable use of SCA data

Health data researchers should use valid selection and analysis methods, apply quality control measures, and be mindful of societal biases being replicated in the data. Only then can research lead to innovations that benefit patients. In most ESCAPE-NET countries the costs of SCA data collection only permit doing this on a relatively small scale,[xiv] with the results representing just a few regions. Therefore, a more effective solution would be to establish national SCA registries that may be used, under certain conditions such as the requirement for consent, for scientific research. A practical option would be for SCA data collection to join existing registries in cardiology, such as the National Heart Registry in the Netherlands.[xv] National SCA registries should include clinical information and ideally also ECG data stored in AEDs. The latter requires establishing an infrastructure for obtaining these data (which becomes easier once AEDs with options for internet connection are more widely used). In the Netherlands, the ARREST group has been the only collector of these life-saving 'AED data', but due to budget constraints this practice was discontinued in July 2021 – which makes the creation of a national registry even more timely. The logistic, ethical and legal specifics should be explored in further (pilot) study but a national registry should at a minimum adhere to governance requirements for quality registration and scientific research, with oversight by an independent entity.[72]

[xiv] Unlocking the potential of SCA data also requires global cooperation. Most research is currently conducted in high-income countries which causes knowledge disparities and a growing digital divide with low- and middle-income countries (LMICs). It is important for local SCA research to be conducted in LMICs, and for researchers such as those in ESCAPE-NET to share their experiences with other parts of the world to promote global health.

[xv] Nederlandse Hart Registratie. [https://nederlandsehartregistratie.nl/]

National registries make research easier by concentrating the data and they may also protect patient rights better by centralising ICT security and ethics oversight. In a centralised system more resources may be available for creating a secure environment, which is increasingly important given growing cybersecurity concerns, and the negative effects of diffused ethics oversight are mitigated. At a minimum, an anonymised dataset of essential characteristics of all SCA victims should be collected, but likely the linkages with other datasets and the desired depth of the information require pseudonymisation of the data. This should be done through a trusted third party that obtains the data from the care setting and then codifies the data. If data that may implicate patients or next-of-kin (e.g. genetic data) are included in such a national registry, a policy and committee for disclosure of findings should be instated.

The most convenient option would be to have SCA registries mandated by law and funded through national governments. However, we learnt from the population registries used in the Second World War and more recently in discussions about tracing apps for COVID-19 that governmental databases pose risks to citizens in situation where democracy is not respected; and that centralised solutions may lack public support.[73] Alternative de-centralised solutions should therefore be explored, e.g. the use of blockchain by independent cooperatives.[74] Any initiative should be accompanied by ELSI guidance and by qualitative research on the non-quantifiable aspects of SCA. Ethical study would be valuable into the effects of classification and on the relation between data and biopower; as well as on practical tools for promoting trust in health data research, in particular for national registries. For instance, storing ethical meta-data that specifies the context of data collection and the conditions of consent, might promote trust in sharing of data with researchers outside the registry.

A national registration system could also include an option for feedback on CPR quality to first-responders, which would be in line with the perceived social responsibility of effectively using an AED when someone is in need. Moreover, national SCA registries might provide the opportunity for researchers to connect with patients in the registry through an online communication infrastructure. Such an infrastructure could help to share the results of studies as a way of thanking participants for their contribution, to be transparent on governance, and to provide more detailed information and consent options. For instance, a variation on 'meta consent'[75] might be used where SCA patients can register whether they are emotionally and cognitively ready to speak with researchers and give consent to more detailed data collection.

Given the importance of maintaining societal trust and the respect for individual persons, I find that research should always require informed consent from SCA survivors if these can be contacted. In some cases, an opt-out option might be sufficient if the research is disproportionately harmed and the individual risks are minimal – but this should be determined for each study separately. Further research is needed on whether control by relatives is desirable. I do think that relatives of deceased SCA

patients should be informed about any research that may have health implications for them. Regarding the consent of SCA survivors, the studies in this thesis suggest that contact should be deferred until the patient is able to provide genuinely informed consent. Cognitive functioning of SCA survivors is returned to normal in most patients when measured through telephone interview after 6-12 months,[76] but my patient interviews suggest that a waiting period of three months may be sufficient. While some SCA researchers collect consent while the patient is in hospital, I find that this should be discouraged or at the minimum be preceded by a test of cognitive capacity. Contacting SCA survivors by telephone before sending consent documentation was appreciated by ARREST participants who said it gives them trust in the researchers, which is echoed in the literature.[77] Individual contact is not only trust-promoting but also helps researchers explain study aims, context, governance structures and potential harms – complex issues that are made more accessible by verbal explanation. Public distrust towards health data sharing is known to diminish when study rationale and potential benefits are explained.[78] Further work is needed to tailor consent procedures to certain populations (e.g. children or people with low levels of literacy), to establish what level of detail in the provided information would be desirable, and regarding how the public can be supported in being responsible trusters.

This thesis also finds that *broad consent* for participation in academic SCA research is ethically justifiable. In addition to views of participants themselves, the main normative arguments are convincing: patients should have the autonomy to give broad consent; and specific consent may lead to selection bias, especially when participants are lost to follow-up or when 'consent fatigue' occurs.[79] The text of the GDPR (Recital 33) is generally acceptive of broad consent, but the guidelines of Working Party 29 and its successor (the European Data Protection Board) note that broad consent is only allowed in exceptional situations where research purposes cannot be fully specified.[80,81] This discrepancy creates uncertainty for researchers and discussion among ethicists and health law experts. It seems that an approach tailored to context is most appropriate (see also Figure 12-2).[82] While specific consent might be more autonomy-enhancing, broad consent is acceptable if specific consent would limit data validity and when data are pseudonymized or not regarded as very sensitive. Moreover, specific consent models may lead to distrust as they come with the 'subtext' that good decisions about whether or not to participate require people to be informed on a detailed level: "Why else should people continuously be informed, be asked and have to decide, if it was not because the *embedded normativity* in this model says it is so important to make this type of informed decision instead of transferring that responsibility to researchers and ethics committees?" (p. 900).[83] In other words, the normativity in specific consent implies that people's privacy is not safeguarded when they do not make the specific decisions themselves. Further research is needed on the relation between trust and the type of consent for health data research.

I should also note that trust is not always voluntary in the medical setting and that it inherently involves being vulnerable. Further research (perhaps through the lens of feminism) is needed on the concept of vulnerability and how this influences decision-making around data-driven research and care in the field of acute and critical illness. Patients with devices like ICDs might be especially vulnerable as there are concerns that these devices might be hacked. Vulnerability also depends on the actual risk of harm related to this type of data-driven research, which should be mapped in future studies. Not only the risks to patients, but also to first-responders require further attention. The fact that every bystander can perform CPR to SCA victims may contribute to interpersonal solidarity but also "requires constant surveillance of everyone by everyone" (p. 95).[56] Research is needed on the privacy and data protection aspects of mobile first responder apps such as those by PulsePoint or HartslagNu.

In my work I found that uncertainty caused by the GDPR and disagreement on underlying values may harm the research enterprise, but that the regulation itself does not. There is a lack of guidelines regarding post-mortem data and for data obtained outside the patient-doctor relationship, such as AED data. Some have suggested that the latter requires a new legal concept for medical secrecy, namely patient secrecy (*'patiëntgeheim'*), but in accordance with evaluations of that proposal I find it unnecessary as the GDPR's principles sufficiently protect patient privacy.[84] Instead, what is needed to ensure that trust in researchers is well-placed, is self-regulation in specific settings. Governance requirements for SCA data research should be specified in a code of conduct devised through an interdisciplinary process. This code can be either sector-specific or consist of adding a clause on the SCA setting to a general code for health data research. Such a code should include provisions related to post-mortem data use. In addition, legal guidance should be available to researchers in order to remove misconceptions about the GDPR, e.g. the idea that the law forms an obstacle for registering ethnicity.[85] Moreover, data management and research integrity courses may help to promote good practices (the majority of data breaches is due to human error)[86] and serve as a reminder for relevant societal values and norms. However, in relation to hard choices and moral dilemmas such as those about equity, codes and courses are of little use. In these situations, deliberation with stakeholders is needed.

12.3.2 Towards data democracy, including after death

I find that a democratic approach to scientific research with health data requires, first of all, oversight by an REC or a similar ethics committee specialised in data research (in addition to legal evaluation by a Data Protection Officer). Assessment by an ethics committee should be done even when data are anonymised or consent is asked, given the limitations of these safeguards and the potential societal impact of (big) data research. Despite the requirement for ethics approval of research protocols in the Helsinki Declaration, this is often interpreted to apply only to intervention research and

there has been an ongoing scholarly debate about the need for ethics committee approval of observational studies.[87-89] Just like biobanking participants have gained more rights over the years, 'data subjects' are slowly moving towards having a similar ontological status as 'human subjects' in research and I find that research institutions' policies should reflect this. At my institution (AMC) this is currently not the case, e.g. the ARREST registry and other critical care databases were not eligible for ethics assessment as they do not address a specific study question.[90] Only on researchers' initiative did ARREST finally received an evaluation, which helped to improve consent policies. In my view, this should be standard practice for health databases.

The composition of RECs should reflect the affected stakeholders and include experts on big data to enable proper evaluation of such studies.[91] Evaluations should be based on context and reason from 'contextual exceptionalism' rather than genetic exceptionalism (see Section 12.1). Moreover, I suggest that REC assessment should be proportional to the context and evaluate at least: the ethical justification of governance strategies in the context of the study; the return of individual findings and whether the right not to know should be upheld; and the societal value and impact of the proposed study. The latter suggestion is debatable and requires further research. Additional study is also needed on how uses of data can be evaluated in light of an Open Science future; on how RECs can be supported in assessing privacy aspects of big data; and about the desirability of harmonising ethics review in international studies. While harmonisation might reduce burdens on researchers,[92] based on my experiences in ESCAPE-NET I tentatively conclude that international laws and guidelines already protect the fundamental rights of patients (although international guidelines about post-mortem data use are needed) and that the specific contextual/cultural details should be left to assessment within a country.

Health data research also requires interdisciplinary collaboration between relevant experts and two-way engagement with the communities from whom the data are derived. For instance, if a project aims to develop smartwatches that alert the EMS dispatch centre in the case of SCA even when there are no bystanders, this is ideally done through collaboration between technical developers, cardiologists, epidemiologists, EMS professionals, psychologists, ethicists and related humanities scholars, legal and financial experts; and in engagement with SCA patient representatives and next-of-kin, as well as the general public. Such multi-stakeholder deliberation is increasingly required by funders as research data governance comes to be viewed as data democracy,[91] which refers to the deliberative approach needed to uphold and improve the social contract regarding health data research. Public and patient engagement (PPE) did not receive enough specific attention in this thesis but I consider it a key part of building trust for health data research, especially for studies surrounded by uncertainty or moral dilemmas.[93] I hypothesise that engagement also promotes integrity, as it prevents researchers taking an the objective stance and helps

them to see data subjects as real people. Further study is needed to see if this is true; and on how to prevent biasing PPE initiatives towards certain groups, especially because the implications of data research may be more serious for minorities. Of note is that not only people in the database should be engaged, but also the public experiencing societal implications of the research, e.g. PPE may help shape shared decision making in the use of big data applications in healthcare (see also Section 12.1.3).

Specific tools for deliberation should be devised which should specifically incorporate the concept of trust, as the value of trust could "tip the scales in favour of one option or another when we must balance conflicting obligations" (p. 85).[40] Such tools could be based on the method of *moral case deliberation* which was devised for the clinical setting but might be helpful for researchers.[94] Another source of inspiration might be the conditions for fair and trust-promoting procedure devised by Norman Daniels in his framework of *accountability for reasonableness* that highlights transparency and publicity of reasons for decisions.[95] Fair process is important as it allows research organisations to pursue their policies with a mandate from society and as it may improve the quality of reasoning and sensitivity to local context. Further work is needed to study how these requirements can be incorporated into a deliberative method for health data research and on what the role of bioethicists should be in this process.

Another topic I want to mention that has not been fully explored in this thesis is the fact that data democracy is especially important for research when data moves to other contexts, as a 'collapse of context' diminishes public trust. This was seen in the much-discussed *care.data* scandal where data were shared with commercial parties. The ESCAPE-NET researchers are currently investigating methods for keeping the database alive after project funding end – long term sustainability of research databases takes hard work and substantial financial investment.[96] Some options include collaboration with commercial parties, but this was not specified in the initial consent and may be unacceptable to patients. In a study about data governance preferences among research participants in a European diabetes consortium project, 44% did not want their data so be shared with commercial companies.[97] Commercial sharing raises various questions about consent (e.g. whether patients should know all the details of licensing agreements and clinicians' personal financial interests) and data ownership (e.g. whether data subject should share in financial benefits). Researchers should learn from best practices and ethics study should address underlying question about whether the concept of the free market is appropriate for health data research:

> "No other mechanism for organizing the production and distribution of goods had proved as successful at generating affluence and prosperity. (…) Today, the logic of buying and selling no longer applies to material goods alone but increasingly governs the whole of life. It is time to ask whether we want to live this way." (p. 5-6)[98]

Lastly, all of the above recommendations also require consideration of deceased subjects' data privacy. As Hannah Arendt said, the public space must transcend the lifespan of mortals.[99] Societal dialogue is needed and ethicists should take a proactive approach in considering this topic, to help avoid potential future scandals related to the misuse of data from the dead, caused by lack of regulation. In addition, death is a tool with which we understand life and studying post-mortem data may give insight in the (minimum) requirements for governing living people's data, or into the environmental concerns raised by storage of big datasets. Further research is needed on the desirability of post-mortem data donation schemes, including the views of people who would not want to partake in such schemes. A question that requires philosophical study is whether the privacy rights of deceased data subjects diminish with time. Such fundamental questions cannot be addressed through the lens of bioethics alone, but require integration with metaphysical theories like the philosophy of information and with epistemology. This broader view would also be useful in addressing challenges around data-driven research like questions around AI, quantum computing, or the next best thing in the 'hype' cycle.[xvi] While studying emerging technologies requires a sensitivity to moral change, the core ethical themes remain familiar.

12.4 Conclusion

In conclusion, this thesis was the first to describe the ethics of big data research in the field of SCA. I have used an empirical ethics approach to provide insight into relevant values and norms, including those of the affected population and their next-of-kin. The studies in this thesis provide specific recommendations for the SCA setting, but my findings also have broader relevance for health data research. I found that there are no *a priori* objections to such research, but that the goodness of science can only be evaluated when ethical aspects are taken into account that relate to the data themselves, to their governance, and to their eventual use. I have argued that context is key: whether you have something to hide, like Anne Frank did, depends on what you are hiding and whom you are hiding it from. Context changes along with societal and technological changes, so ongoing ethical questioning and inclusive qualitative research are needed. In the age of big data, even deceased patients can have privacy concerns. This thesis provides inspiration for various themes to be explored in further collaborative work, such as the sharing of data in public-private partnerships. Time magazine said that *you* control the information age,[xvii] but really it is up to *us* to shape the governance of health data research together.

[xvi] Gartner's hype cycle graphically represents the adoption and social application of new technologies. [https://www.gartner.com/en/research/methodologies/gartner-hype-cycle]

[xvii] Reuters, 20 January 2007. 'Time magazine's "Person of the Year" is You'.