Precision Health Approaches: Ethical considerations for health data processing

Inaugural dissertation

to

be awarded the degree of Dr. sc. med

presented at

the Faculty of Medicine

of the University of Basel

by

Lester Darryl Geneviève

From Quatre-Bornes, Mauritius

Basel, 2023

Original document stored on the publication server of the University of Basel **edoc.unibas.ch**

This work is licensed under a Creative Commons Attribution 4.0 International License.

Approved by the Faculty of Medicine

On application of

First Supervisor: Prof. Dr. med. Bernice Simone Elger

Second Supervisor: PD. Dr. Tenzin Wangmo

Further Advisor: Prof. Dr. med. Thomas Perneger

External Expert: Prof. Dr. med. Bettina Borisch

Basel, 10 May 2021

Prof. Dr. Primo Leo Schär

Dean

Table of Contents

Acknowledgments

I would like to express my appreciation to my PhD supervisors, Prof. Bernice Elger and Dr. Tenzin Wangmo, and to my further advisor, Prof. Thomas Perneger. I am deeply indebted to Prof. Elger for her continuous support and guidance during the conduct of our project within the framework of the National Research Programme 74 (NRP-74) on Smarter Health Care, and for offering me the academic freedom necessary to pursue my own research ideas and interests. Dr. Wangmo has been a great mentor and colleague since the beginning of my doctoral journey, helping me through tough times that every student faces and always pushing me past my limits. Prof. Perneger has been of a great support through his thoroughness and accuracy in evaluating my PhD work, which allowed me to be more critical and thoughtful about my own research. I am also very grateful to Prof. Bettina Borisch for agreeing to be my external expert.

I would also like to thank the Swiss National Science Foundation for funding our NRP-74 project and my PhD work. I am grateful to the Swiss expert stakeholders for their participation to the interviews conducted for this project. Concerning the smooth running of my work at the Institute for Biomedical Ethics, a special thanks goes to Mrs. Anne-Christine Loschnigg, her help was deeply appreciated during these four intense years.

I would also like to thank all my colleagues and friends at the Institute for Biomedical Ethics for the interesting discussions and support. Andrea Martani has been an amazing colleague and friend, without whom many aspects of this project would not have been achievable. Thank you also for the numerous chess games that we played, which allowed me to blow off some steam after work. My special thanks also goes to Maddalena Favaretto, whose support and friendship during the ups and downs of this doctoral journey were invaluable. Thank you Michael Rost for your friendship and for always taking the time to share the latest information and webinars relevant to my research interests.

Last but not least, I would like to express my deepest gratitude to my family. First to my wife, Christina, whose love, continuous support and unwavering faith have allowed me to see this PhD through to the end. Second, this accomplishment would also not have been possible without the support and love from my parents Emmanuelle and Clifford, my sisters and brother. Thank you Pedro Barrio for taking care of the family in Geneva, which I was unfortunately not able to do as often as I would have liked from Basel. A special thought to my late stepfather, Gregory Zurcher, who had always believed in me and to whom I dedicate this PhD.

Summary

This thesis provides insights and recommendations on some of the most crucial elements necessary for an effective, legally and ethically sound implementation of precision health approaches in the Swiss context (and beyond), specifically for precision medicine and precision public health. In this regard, this thesis recognizes the centrality of data in these two abovementioned domains, and the ethical and scientific imperative of ensuring the widespread and responsible sharing of high quality health data between the numerous stakeholders involved in healthcare, public health and associated research domains. It also recognizes the need to protect not only the interests of data subjects but also those of data processors. Indeed, it is only through a comprehensive assessment of the needs and expectations of each and every one regarding data sharing activities that sustainable solutions to known ethical and scientific conundrums can be devised and implemented. In addition, the included chapters in this thesis emphasize recommending solutions that could be convincingly applied to real world problems, with the ultimate objective of having a concrete impact on clinical and public health practice and policies, including research activities. Indeed, the strengths of this thesis reside in a careful and in-depth interdisciplinary assessment of the different issues at stake in precision health approaches, with the elaboration of the least disruptive solutions (as far as possible) and recommendations for an easy evaluation and subsequent adoption by relevant stakeholders active in these two domains.

This thesis has three main objectives, namely (i) to investigate and identify factors influencing the processing of health data in the Swiss context and suggest some potential solutions and recommendations. A better understanding of these factors is paramount for an effective implementation of precision health approaches given their strong dependence on high quality and easily accessible health datasets; (ii) to identify and explore the ethical, legal and social issues (ELSI) of innovative participatory disease surveillance systems – also falling under precision health approaches – and how research ethics are coping within this field. In addition, this thesis aims to strengthen the ethical approaches currently used to cater for these ELSIs by providing a robust ethical framework; and lastly, (iii) to investigate how precision health approaches might not be able to achieve their social justice and health equity goals, if the impact of structural racism on these initiatives is not given due consideration. After a careful assessment, this thesis provides recommendations and potential actions that could help these precision health approaches adhere to their social justice and health equity goals.

This thesis has investigated these three main objectives using both empirical and theoretical research methods. The empirical branch consists of systematic and scoping reviews, both adhering to the PRISMA[1] guidelines, and two interview-based studies carried out with Swiss expert stakeholders. The theoretical branch consists of three chapters, each addressing important aspects concerning precision health approaches.

Thesis outline

This thesis is composed of the following chapters.

Chapter 1: This chapter provides an introduction to essential elements forming the foundation for the remaining chapters of this thesis. It covers the broad topic of health data and their specific properties that make them different from industrial data with reference to their high complexity and information density. It also highlights the importance of having both *Big Data* and *Small Data* approaches for precision health approaches, and the often-necessary use of artificial intelligence tools (such as machine learning) for *Big Data* experiments to make sense out of *Big Data*. In addition, it provides an in-depth overview of two precision health approaches, namely precision medicine and precision public health, and highlights their ethical challenges.

Chapter 2: This chapter provides the overall research objectives of this thesis and explained the different methodologies used to answer them in the chapters 3 to 9. The different steps of the research process for this PhD work are explained. The author contributions are also discussed.

Chapter 3: This chapter investigates and compares the different types of barriers and facilitators influencing the processing of health data in Danish and Swiss projects through a systematic review of the literature. Success mechanisms in Denmark are also analyzed and their potential implementation in the Swiss context assessed.

Chapter 4: This chapter investigates, through an interview-based study, systemic factors influencing the fair sharing of health data from the perspectives of Swiss expert stakeholders. For this chapter, a theoretical thematic analysis was carried out to identify themes and subthemes pertinent to the objectives of this study.

Chapter 5: This chapter also investigates, through an interview-based study, the individual notions of fair data sharing from the perspectives of Swiss expert stakeholders. A theoretical thematic analysis was also performed to identify themes and subthemes that hinted to the

[1] Acronym for: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

fairness of the data access negotiation process between the original data collectors and data recipients. A distributive justice lens was used to provide the overall analytical framework for this study.

Chapter 6: This chapter investigates, through a systematic scoping review, how different country-specific platforms of the Influenzanet consortium – a participatory syndromic surveillance system for influenza-like illness – have implemented research ethics components. It then highlights the need and suggests ways to harmonize ethical approaches to the ELSIs identified, in order to prevent future obstacles to collaboration opportunities between the different platforms with regard to the evolving data protection landscape at the European level.

Chapter 7: This theoretical chapter investigates in-depth the different ELSIs identified in the disease surveillance platforms of the Influenzanet Consortium, and provides recommendations to strengthen their ethical approaches. Different components are addressed such as the electronic consent procedures, the need to protect the privacy of participants, the need to cater for justice issues and lastly, the need for research ethics committees to develop the necessary expertise and skills to tackle the novel ethical challenges brought by such innovative surveillance methods.

Chapter 8: This theoretical chapter investigates how precision public health approaches, being considered as an additional tool to fight the ongoing COVID-19 pandemic, could be influenced by structural racism. After an in-depth analysis, recommendations are made to help alleviate the impact of the latter to ensure that racial and ethnic minorities are not discriminated in the public health response.

Chapter 9: This debate chapter investigates the impact of structural racism on the data ecosystem of precision medicine initiatives, and forecasts that its deleterious effects, in terms of suboptimal access, low quality of care and biased deliverables, are going to be reinforced over time due to the iterative data exchange occurring between research and clinical sectors through learning healthcare systems. Thereafter, it provides some potential actions and recommendations to mitigate the impact of structural racism at three specific nodes of the data cycle in precision medicine initiatives.

Chapter 10: A general discussion, informed by the previous chapters, addresses the findings and practical/policy implications of the empirical and theoretical branches of this thesis in chapter 10. The general limitations of this thesis, its implications for future research and conclusions are also provided at the end of this chapter.

Chapter 1: Introduction

“In God we trust, all others must bring data” – W. Edwards Deming

1.1 Health data: etymology, definition and their specific properties

The word, *data*, has its etymological roots from the Latin word, *datum*, meaning "a fact given or granted", which dates back to the 1640s [1]. It is only in the 1940s that the term got its modern connotation – with its relation to the computer science field made explicit – that is, "transmissible and storable information by which computer operations are performed" [1]. Subsequent terms such as 'data-processing', 'database' and 'data-entry' only came into use in 1954, 1962 and 1970 respectively [1]. Today, data is defined as, "information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer" [2]. In this regard, since the start of the third industrial revolution in 1969 [3] – aligned with the *Information Age* [4] – digital data have become increasingly pervasive in every sector of our contemporary societies. However, the digitalization wave has had different sector-specific penetration rates. Indeed, while many sectors have moved into the fourth industrial revolution, with the large scale deployment and application of artificial intelligence (AI) tools, the healthcare sector is still held in the third one, struggling lately with the widespread adoption of electronic health records (EHR) and the early contemplation of AI tools to guide clinical decision-making [5, 6]. Before proceeding to the specific characteristics of health data, it is important to define precisely the latter. On this matter, Recital 35 of the General Data Protection Regulation (GDPR) defines personal data concerning health as:

> "all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject. This includes information about the natural person collected in the course of the registration for, or the provision of, health care services as referred to in Directive 2011/24/EU of the European Parliament and of the Council to that natural person; a number, symbol or particular assigned to a natural person to uniquely identify the natural person for health purposes; information derived from the testing or examination of a body part or bodily substance, including from genetic data and biological samples; and any information on, for example, a disease, disability, disease risk, medical history, clinical treatment or the physiological or biomedical state of the data subject independent of its source, for example from a physician or other health professional, a hospital, a medical device or an in vitro diagnostic test" [7].

In comparison to industrial data, health data sources are often kept in silos, with high information density and complexity, and not easily accessible and shareable due to various

constraints (e.g., technical and political, etc.), including a heightened legal scrutiny aimed at protecting the privacy of data subjects due to their sensitive nature [8, 9]. Additionally, Shilo, Rossman and Segal [10] described seven axes of health or healthcare data, i.e. specific quantitative properties with different computational and processing abilities. These axes of any health data source include (i) the size of the sample available (i.e. the number of participants), whose determinants comprise organizational and financial considerations (some of these are discussed in chapters 3 and 5); (ii) the extent to which the phenotyping process has been carried out in the dataset, which can include diverse data types such as molecular data, socio-economic and demographics data; (iii) the amount and type of data collected at different time points during the follow-up period of participants to investigate how health parameters evolved over time, which can help in understanding the pathogenesis of diseases; (iv) identifying social interactions between data subjects could provide insights on mechanisms underpinning the development of certain diseases, such as by comparing with relatives' data to better understand gene-environment interactions; (v) having harmonized and standardized data, that would allow the analysis of different datasets simultaneously to answer important research questions, whilst at the same time, fostering research collaborations (these problematics are discussed in chapter 3); (vi) the ability to link datasets to gather information on a specific participant across different databases, such as with the help of a unique personal identifier (also discussed in chapter 3); (vii) the representativeness and generalizability of a health dataset with regard to the population at large (e.g., underrepresentation of racial/ethnic minority groups is a major concern [11], which is discussed in chapters 8 and 9) [10]. Each of these seven axes of health data has its own set of challenges, and it is therefore necessary to have some trade-offs between them (e.g. the depth of phenotyping is lower for EHRs than for biobanks because the latter usually have less participants) [10].

It is also worth noting that the global amount of healthcare data generated for 2020 is projected to be 2314 exabytes [12]. In this perspective, the healthcare and public health domains are gradually turning into data rich digital environments, with the potential to provide better insights on the causality of diseases, including improved, timely, and tailored care and disease prevention strategies to individual patients and to the population in general [10, 13]. Such advancements in healthcare delivery were partly made possible due to some relatively recent legislative reforms backing up the uptake of new health information technologies (e.g. EHRs). These were adopted in the interest of public health, whilst improving the cost-effectiveness of healthcare institutions. In the United States, this is exemplified by the *Health Information Technology for Economic and Clinical Health* (HITECH) Act [14, 15] and in Switzerland, by

the *Bundesgesetz über das elektronische Patientendossier*[2] (EPDG) [16, 17]. Moreover, there are other digital sources of health-related data, including contributions from other sectors (e.g., social media, web-search queries, wearable activity trackers, etc.), which are advancing healthcare and public health [18-20].

Regardless of the source of the health data considered, the analysis of these datasets can be used to either generate transportable knowledge that could potentially benefit other people and society at large, through so-called *Big Data approaches*, or they can be used to predominantly and directly benefit the health of the individuals contributing the data, i.e. through *Small Data approaches* [21]. These two concepts are discussed in the following sections.

1.2 Big and Small Data for Precision Health Approaches

Big Data in Healthcare and Public Health

The concept of *Big Data* was first attributed three defining characteristics by Laney in 2001 (as cited in [22]). These characteristics were specifically in terms of *volume* (i.e. massive datasets), *velocity* (i.e. data generated in real-time or at high speed) and *variety* (diverse data formats: structured, semi-structured and unstructured), commonly referred nowadays to as the three Vs of *Big Data* [22]. Since then, the terminology of *Big Data* has evolved, with the addition of new defining characteristics such as *variability*, *veracity,* and *value* to name a few [23-25]. However, as argued by Kitchin and McArdle [22], the concept of *Big Data* still lacks conceptual clarity. In this regard, the authors explored and analyzed twenty-six datasets, considered to be *Big Data*, for known defining characteristics. They determined that two of these allegedly qualifying characteristics, notably *volume* and *variety*, are not essential elements that datasets need to possess in order to be considered as *Big Data*. In addition, the terminology itself was argued to be misleading, influencing the viewing of *Big Data* mostly in terms of *volume*, and thereby propagating definitional and reductionist misconceptions of what really constitutes *Big Data* [22]. Furthermore, based on their analysis, the authors concluded that two other characteristics, namely *velocity* and *exhaustivity* (i.e. capturing all elements rather than a sample) were the most important distinguishing features that separate *Big Data* from *Small Data* [22]. The exclusion of the *volume* of datasets as a qualifying and integral characteristic of *Big Data* raises the questions of whether certain relatively smaller healthcare and public health datasets, in comparison to industrial standards, can therefore be considered as being part of the *Big Data* continuum due to their inherent complexity.

[2] German for: Federal Law on the Electronic Patient Record

In this regard, the healthcare domain has indeed become a data-rich environment, fueled by technological advances (e.g. next-generation sequencing [26] or any high-throughput data capturing and analysis methods [27]) allowing the integration of omics-data such as, amongst others, genomics, epigenomics, transcriptomics, metabolomics, and proteomics for more tailored and personalized healthcare interventions [28]. For instance, the integration of genomics data to EHRs is now being observed [29]. One example is the University of Pennsylvania's *PennChart Genomics Initiative* that has been able to successfully integrate both structured and unstructured genetic data to EHRs, while contributing to the optimization of clinical decision support tools [30]. Moreover, other routinely collected healthcare datasets are also being used for educational purposes and to advance industrial and academic research. One notable example is the freely accessible intensive care mono-center database, MIMIC-III[3] from Boston's *Beth Israel Deaconess Medical Center* in the Massachusetts [31]. Building on the successes of previous versions (MIMIC and MIMIC-II), MIMIC-III provides detailed and granular information on individual care (e.g. billing and administrative information, medical interventions practiced, laboratory results, hospital discharge summaries, physiological data and free-text data, to name a few) by pooling data from EHR databases, intensive care information systems records and death information from the master file of Social Security Administration [31]. In this regard, MIMIC III could be considered as *Big Data*, relying on some advanced data analytics, including the involvement of both data scientists and clinicians, to make sense out of these datasets in order to potentially predict certain clinical outcomes and thereafter, optimize the effectiveness of interventions [32]. Indeed, the conditions necessary for some health datasets (either clinical or research) to be considered as *Big Data* resides also in the difficulty or inability of health professionals or researchers to analyze, understand and translate those findings to the bedside of patients through traditional data analytical methods [6].

Similar to the advent of *Big Data* in healthcare, the public health domain has also witnessed a rapid and exponential growth of datasets of public health importance due to the previously described digitalization wave and technological developments. In this regard, Mooney and Pejaver [33] described five main *Big Data* streams for public health practice and research by considering participants' context, their biological indicators and administrative medical pathway, and their self-generated data either through devices that automatically record health-related data or through their own normal activities in the digital world (e.g. social media and

[3] Acronym for: Medical Information Mart for Intensive Care

web-based searches). These five sources of *Big Data* are classified as follows: (i) biological or omics (e.g. metabolomics), (ii) geospatial data, (iii) EHR, (iv) personal monitoring data (e.g. wearable activity trackers) and (v) effluent data (e.g. Twitter, Reddit or Google searches) [33]. Although each data source faces its own set of technical limitations [33], some of them have successfully been used or have shown promising results in public health surveillance [34, 35], etiologic research, clinical research [36], and screening (e.g. BRCA1[4] and BRCA2 screening for preventive measures against ovarian and breast cancer in women [37, 38]) [33].

Small Data in Healthcare and Public Health

In contrast to the goal of a *Big Data* approach, which is the generation of transportable knowledge on a specific phenomenon (in terms of enhanced descriptive and predictive capabilities) that can be applied to a wider context, a *Small Data* approach aims to "achieve improved individual-level description, prediction and, ultimately, control for that specific unit" [21]. Therefore, one of the key differences between *Big Data* and *Small Data* resides in the fact that for a *Small Data* approach, the insights generated from data analysis are being used primarily for that specific unit from which the data was collected. However, *Small Data* do not imply that the volume of datasets is small but rather that the purpose of the data collection from a single unit (e.g. an individual, a community, a hospital or clinic, to name a few) is to provide idiosyncratic insights on that specific unit, which a *Big Data* approach is not able to accomplish [21, 39].

Therefore, a *Small Data* approach offers three practical advantages, namely, it allows evaluation criteria for success, in terms of enhanced description, prescription and control, to be aligned with the objectives put forward by the single unit; it generates timely knowledge that can be implemented rapidly in practice, and lastly, it generates transportable knowledge after the identification of "clusters of actionable insights" across individuals (i.e. crucial mechanisms underpinning the success observed in the single unit) [21]. A few examples using a *Small Data* approach include: (i) *MyBehavior* [40], which is an automated mobile app that uses mobile data and subsequently provides personalized and effective health recommendations to its users based on their mobility status in order to increase the burning of calories; and (ii) *TummyTrials* [41], another mobile app that allows self-experimentation in order to help its users identify individualized food triggers for inflammatory bowel disease.

While a *Big Data* approach seems to be well suited for certain health problems where their underlying mechanisms and causal determinants are adequately known, its use is limited in

[4] Acronym for: BReast CAncer gene 1

health problems that are expressed individually, dynamically and with multiple causal determinants. In this regard, the inadequacy of a *Big Data* approach resides not only in the epistemological limitations in finding the appropriate data types to be included in the analysis but also in the fact that important idiosyncratic determinants of health, i.e. those that are specific and meaningful to an individual or a unit, could simply be ignored or excluded by a *Big Data* approach [21]. In this regard, the core of such limitations reflects the typical tension between two main healthcare movements, which are *patient-centered care* (favoring a *Small Data* approach) and *evidence-based medicine* (favoring a *Big Data* approach) [42].

On this matter, important work has already been done to try solving this conundrum, in particular how *Big* and *Small Data* approaches could be integrated for the improvement of the health of individuals and the population [42-44]. Although receiving much attention in recent decades, *Big Data* will not displace *Small Data* but it will be rather a complementary approach or an addition to *Small Data* for improving the health of individuals and the population [21, 42]. Undeniably, it has been argued by Hekler et al. that *Big Data* would provide the big picture pertaining to disease causation by identifying potential causal determinants of health or ill-health and solutions that could be implemented, whereas a *Small Data* approach would allow to go more in-depth in understanding what specific causal determinants and solutions are adapted to the specific unit [21]. In contrast, Sacristan and Dilla have argued that *Small Data* and *Big Data* will both play a role in the implementation of learning healthcare systems, through the assimilation of both *evidence-based medicine* and *patient-centered care*, which will be facilitated through the widespread adoption of electronic medical records (EMR) [42]. Indeed, the authors argued that in the context of a learning healthcare system, *Small Data* will be the foundation of *Big Data* [42] rather than being a complementary approach like Hekler et al suggested [21]. However, for *Small Data* to become progressively *Big Data*, it would require that limitations associated with *Small Data* are addressed (e.g. the lack of systematic collection for research purposes that would allow generalizable knowledge to be derived for other patients). That is, the resulting *Big Data* should be used for medical decision-making purposes and care to improve the health of individuals [42]. Whatever the approach adopted to integrate *Small* and *Big Data* in the future, the former is here to stay and it is even more evident if one looks at rare diseases where *Big Data* simply do not exist [45].

Relevance and Importance of Big Data for precision health approaches

The number of *Big Data* sources for precision health approaches is increasing exponentially, along with the advent of new technologies – such as body sensors and *Big Data* analytics – that allow a more in-depth, real-time and continuous monitoring of the health determinants of

patients and populations. These offer new learning opportunities for the healthcare and public health systems, whilst facilitating a paradigm shift towards systems that will be more proactively and efficiently tackling health and public health threats [46].

First, the current applications of *Big Data* in precision medicine (PM) are mainly at three distinct levels: (i) clinical research, (ii) basic research and (iii) clinical practice [47]. Interestingly, the analysis of *Big Data* is considered to be more prone to the generation of new hypotheses rather than being hypothesis-driven, and by its very nature, it is likely to uncover novel and unsuspected disease mechanisms that would benefit PM initiatives [8]. For instance, *Big Data* experiments could help PM's endeavors in understanding the pathogenesis of certain diseases by analyzing the already existing clinical data silos held in hospitals, if the latter could be pooled and linked with other datasets (e.g. genetic data) for analysis. Indeed, these clinical data silos are often managed by IT[5] systems dating back to an era where such new methodological approaches and the need to combine and share those data were not foreseeable [8]. Second, according to a review by Dolley [48], the applications of *Big Data* in precision public health (PPH) currently includes four broad categories or sectors, which are (i) public health surveillance for diseases and signal detection (e.g. using Twitter data for COVID-19 [49]), (ii) prediction of risks (e.g. for antimicrobial resistance [50]), (iii) tailored public health interventions for homogeneous sub-groups of the general population (e.g. personalized vaccination, also called "vaccinomics", that identifies genetic factors influencing individual responses to maximize the effectiveness of vaccines administered [51]), and lastly, (iv) a better comprehension of diseases (e.g. through etiologic research) [48]. Before proceeding to the concepts of PM and PPH, which are discussed in details in section 1.4, it is important to understand how insights are derived from these big health datasets. This is explained in the following section.

1.3 Big Data: Artificial Intelligence versus Limits of the Human Mind

The term "Artificial intelligence" (AI) was coined in 1956, following a two-month workshop held at Dartmouth College in the United States by Stanford renowned computer scientist, John McCarthy with the help of colleagues from, amongst others, Princeton, Massachusetts Institute of Technology and IBM[6]. There have been numerous candidate definitions for AI, based on its two main goals. For instance, whether AI wants to adopt a human-centered approach, i.e. mimicking human performance, or a rationalist approach, where systems will be able to think

[5] Acronym for: Information Technology
[6] Acronym for: International Business Machines Corporation

or act rationally [52]. In spite of not having a clear consensus on what clearly constitutes AI [53], its applications have mushroomed over the last few decades, owing to the ongoing increasing computational power, with global market revenues estimated at 7,35 billion US dollars in 2018 [54].

Recently, a comprehensive definition of AI has been put forward by the *high-level expert group on Artificial Intelligence* of the European Commission, namely that AI systems are, "software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions" [55].

Our contemporary societies have witnessed the widespread implementation of AI technologies in different sectors, such as in the gaming industry, speech recognition applications, robotics or even spam fighting, just to name a few examples [52]. The healthcare sector, becoming increasingly digitalized and information rich (e.g., with the widespread implementation of EHRs), is also falling under the realm of *Big Data*, and subsequently AI tools are being considered to make sense out of such exponential growth of datasets in order to improve the overall efficiency of the care system [56, 57]. Such massive and complex datasets can no longer be managed and understood by humans, without the help of *Big Data* analytics. In that regard, Topol's review forecasts that AI technologies will become a ubiquitous apparel for all physicians [56]. The review also reveals that the potential implementation of AI in medical specialties (e.g., radiology, gastroenterology, ophthalmology, etc.) has indeed proved to be successful or at least, the results looked promising [56]. Therefore, in the very near future, AI is likely to be the norm in the healthcare domain, in particular given that its financial sustainability is not being threatened.

Indeed, following a financial analysis on the use of AI in healthcare, the company *Accenture* stated that, "AI thinks and pays for itself", estimating annual savings of up to 150 billion US dollars on healthcare costs in the US by 2026 [58]. Similarly, the public health sector is not immune to the transformational change brought by the implementation of AI technologies. Likewise, AI has proven to be key in understanding complex, diversified, novel but valuable data streams (e.g., Twitter, smartphone sensors, meteorological data) to guide public health

practice [59, 60]. In this regard, one of the most important set of AI technologies used in the healthcare and public health domains is machine learning.

Machine learning techniques can be differentiated into the three most commonly used classes, namely (i) *unsupervised*, (ii) *semi-supervised* and (iii) *supervised* machine learning techniques [28, 55]. In *unsupervised* machine learning, the input dataset used to train the classifier is not labeled. Instead, the machine learning algorithm learns to identify clusters or patterns that are hidden in the training dataset and agglomerates the data into meaningful clusters. This method is often used in oncology for molecular subtyping [28]. For instance, one example is RAPID[7], an unsupervised machine learning algorithm that has been used to identify risk stratifying glioblastoma tumor cells [61]. In *semi-supervised* machine learning, the input dataset used to train the classifier contains both labeled and unlabeled data, that allows the classifier to first understand the data structure, and then make predictions for the unlabeled data. In *supervised* machine learning, the training dataset is fully labeled, and therefore allows the learning model to subsequently maximize its performance in terms of accurately predicting or mapping new data samples according to the labels used [28]. In spite of their potential to improve healthcare and public health outcomes, one inherent issue arising from the use of many machine learning algorithms resides in the fact that their decision processes cannot be explained by their developers, a problem commonly known as the "black box" [62].

This lack of understanding on how some machine learning algorithms reach certain decisions creates some reluctance in healthcare professionals planning to use them in their routine work and regulatory bodies to allow their implementation for medical care [63, 64]. In this regard, it was argued that, "[p]hysicians who use machine-learning systems can become more educated about their construction, the data sets they are built on, and their limitations. Remaining ignorant about the construction of machine-learning systems or allowing them to be constructed as black boxes could lead to ethically problematic outcomes" [65]. Therefore, Watson and colleagues have argued for a collaborative enterprise – involving physicians, data scientists and patients – in order to develop novel methodologies that could explain decisions made by machine learning algorithms that are pertinent to the model itself, i.e. model-centric, and to the patient, i.e. subject-centric explanations [62]. Such an endeavor is important, given that machine learning algorithms are expected to make significant contributions in the nascent but

[7] Acronym for: Risk Assessment Population Identification

rapidly developing fields of PM and PPH [8, 21, 28, 66, 67]. These two emerging concepts are discussed in the following sections.

1.4 Precision Medicine and Precision Public Health

Precision Medicine

PM is defined as, "an approach to disease treatment and prevention that seeks to maximize effectiveness by taking into account individual variability in genes, environment, and lifestyle" [68]. Importantly, the ultimate objective of PM is to allow individual patients to obtain the most effective treatment at the right dosage and time, with the least possible side effects [69]. Such an endeavor is supported by technological advances (e.g. cheaper tumor profiling or genomic sequencing) allowing the integration of state-of-the-art molecular analysis techniques (e.g. multi-omics) for profiling diseases, including the inclusion of other factors that shape an individual's health such as, amongst others, environmental exposures [69-72]. It is interesting to note that, in spite of the diverse aims of PM in terms of diagnosis, disease prevention and treatment, a greater emphasis has been put towards so-called "precision pharmacogenetics" and the identification of a patient's genetic markers for the most suitable drug adapted to his or her condition [67]. It is also important to highlight that PM is not revolutionary *per se* but rather follows a natural evolutionary process from more traditional approaches to medicine, clinical trials and care [72]. Additionally, PM will induce a paradigm shift, from the traditional dichotomy observed between institutions dedicated to clinical care and those dedicated to biomedical research, to another key concept for PM to thrive, known as *learning healthcare systems*, where the boundaries between clinical care and research are almost non-existent [73]. This transition is already occurring on a pilot level, with the launch of some NIH[8]-funded initiatives such as *Implementing Genomics in Practice* (IGNITE) and *Electronic Medical Records and Genomics* (eMERGE) [73].

In Switzerland, there are two main complementary PM initiatives, specifically the *Swiss Personalized Health Network* (SPHN) [74] and the *Personalized Health and Related Technologies* (PHRT) [75]. The SPHN was initiated following a mandate to the *Swiss Academy of Medical Sciences* in 2016 by the *Federal Office of Public Health* (FOPH) and the *State Secretariat for Education, Research, and Innovation* (SERI) [74]. Recognizing the limitations of the current Swiss health data landscape to sustain the implementation of PM, the main objective of the SPHN between 2017 to 2020 was to lay the foundations for a national data infrastructure that would allow the sharing of health-related data in an interoperable manner for

[8] Acronym for: National Institutes of Health

research purposes between participating partners (e.g. University hospitals), in order to advance personalized health and personalized medicine in Switzerland [76]. A lot of driver projects have been implemented in this regard, such as the *Swiss Personalized Oncology* (SPO) project [77]. In comparison, the PHRT is part of the Swiss Federal Institutes of Technology (e.g. ETH Zürich and EPFL in Lausanne) and it aims to support the implementation of PM by providing cutting-edge technologies that would allow not only the integration of (multi)omics data in the healthcare IT infrastructure but these new technologies would also support healthcare providers during clinical decision-making [78].

In France, there are also two main PM initiatives, namely the *French Society for Predictive and Personalized Medicine*[9] (SFMPP) [79] and the *French Plan for Genomic Medicine 2025* (*FPGM*)[10] [80]. Starting in 2015, the FPGM has a 10-year plan to allow (i) the uptake of genomic sequencing into routine clinical practice, with special attention to oncology patients and those suffering from rare diseases, where the resulting data (both clinical and genomic) will be stored in a national database, and (ii) to establish a nationwide genomic sector, dedicated to the advancement of genomic medicine in industrial and medical settings [81].

In the UK, there were also initially two main PM initiatives: (i) *Genomics England* [82], and (ii) *Precision Medicine Catapult* (PMC). Unfortunately, the latter was closed prematurely before its launch [83]. In contrast, *Genomics England*, which started in 2013, is still active and includes the UK's flagship PM enterprise, the *100,000 Genomes Project*. The latter aims at paving the way to PM approaches by sequencing 100,000 genomes – on top of collecting data (EHR, phenotypic and genomic) from 70,000 individuals suffering from either cancer or rare diseases – whilst promoting capacity-building for the genomic industry [81, 84].

Other PM initiatives exist throughout Europe, and they include, amongst others, *FinnGen* in Finland [85], *Italian Society for Personalized Medicine* (SIMeP) [86] and the *Estonian Genome Project* [87].

In the United States one of the most prominent PM initiatives, funded by the National Institutes of Health, is the *All of Us* research program, which aims to collect data on at least one million people from the US [88]. Other US initiatives include, amongst others, the *California Initiative to Advance Precision Medicine* (CIAPM) and University of Columbia's precision medicine initiative [89].

[9] In French: Société Française de Médecine Prédictive et Personnalisée
[10] In French: Plan France Médecine Génomique 2025

In Canada, there is only one national organization dedicated to genomics: *Genome Canada* [90]. Through its twenty years of existence, *Genome Canada* has been trying to exploit the potential benefits that could be derived from genomics to implement precision approaches to health. However, the activities of *Genome Canada* are not limited to health since the organization also explores other avenues where the field of genomics can have an impact (e.g. in agriculture) [91]. Other PM initiatives also exist at the provincial level such as the *Ontario Personalized Medicine Network* (OPMN) [92] and *Génome Québec* [93].

In Asia, there are also numerous PM initiatives being implemented, such as the recently launched *Genome India Project* in 2020 [94], another project from the Japanese National Cancer Center known as *TOP-GEAR*[11] [95] and the *China Precision Medicine Initiative* [96] to name a few. In Africa, there is a multi-country project namely, *Human Heredity and Health in Africa* (H3Africa) [97], which is dedicated to advance the implementation of PM for African populations.

Precision Public Health

It is important to understand that PM can improve the health of an individual by tailoring interventions to his or her specific needs, but such interventions do not inevitably result in standardized health benefits for all subgroups of the population [67, 98]. For example, racial and ethnic minorities could be left behind in reaping the health benefits if PM does not account for societal (e.g. structural racism) or environmental challenges [67, 99]. In this regard, PPH can help alleviate some of these issues.

PPH can be defined as, "the application and combination of new and existing technologies, which more precisely describe and analyse individuals and their environment over the life course, to tailor preventive interventions for at-risk groups and improve the overall health of the population" [100]. In essence, PPH is similar to PM but rather than focusing on an individual, PPH ensures that a specific population receives an intervention tailored to their needs in a timely manner, in order to maximize the effectiveness of such an intervention [101].

In a 2020 review, Khoury et al [102] provided interesting examples of the application of PPH approaches in real-world settings, relying on *Big Data* approaches, which include: (i) *public health surveillance* (that is, the "ongoing, systematic collection, analysis and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice" [103]), (ii) *implementation science in the public health arena* (that is, the science

[11] Acronym for: Trial of Onco-Panel for Gene-profiling to Estimate both Adverse events and Response

promoting the timely uptake of new public health research findings and evidence to safeguard and promote the health of populations [104]), (iii) *data science* for advanced predictive analytics, and lastly (iv) *pathogen genomics* (for instance, the use of next-generation genomic sequencing for pathogen characterization allows an early detection of disease outbreaks, an improved surveillance and a subsequent adapted public health response, in particular for transmissible diseases by providing insights on disease transmission mechanisms [105]) [102].

Participatory Disease Surveillance: another facet of PPH

Richard Horton, the editor-in-chief of the *Lancet*, described PPH as, "using the best available data to target more effectively and efficiently interventions of all kinds to those most in need" [13]. In this regard, it can be argued that participatory disease surveillance systems are also part of the PPH spectrum. The argument for its inclusion is explained in the following paragraph.

Just like advances in information technology have contributed to the phenomenon of *Big Data*, they have also empowered individuals to participate more actively in public health activities (such as in public health surveillance) [106, 107], by sharing not only health-related information via web-based and mobile platforms (e.g. reporting symptoms of a specific disease such as influenza and geospatial data for detecting outbreak location) but by also receiving in return timely information that would allow the users to minimize risky behaviors and subsequently reduce the potential impact of the disease [108]. Two of the numerous advantages are: (i) they allow an early reporting of disease outbreaks by bypassing the different traditional public health infrastructural levels that information need to go through for conventional approaches [109], and (ii) they allow a better geospatial detection of outbreak locations for targeted interventions [110]. Therefore, it can be argued that participatory disease surveillance systems also contribute to more precise, timely and preventive public health interventions, aligned with the definition provided by Horton earlier [13].

Centrality of data and obstacles to its processing for precision health approaches

The reliance of precision health approaches on data cannot be emphasized enough [67]. Indeed, it is crucial to highlight that PM and PPH are also struggling for implementation because of barriers hindering the effective exploitation of health data (e.g. difficulties to link health data collected from different sources, or semantic issues leading to ambiguous terminologies to name a few) [67]. For instance, it was highlighted that for PM initiatives to function as they should, several criteria needed to be fulfilled. These are: (i) numerous and different data types from the patient should be readily accessible, (ii) these data should be analyzed and integrated in such a way that they provide meaningful results that can be implemented in clinical practice,

(iii) tackling logistical barriers that hinder the delivery of recommendations to patients, and lastly (iv) providers and payers should accept the entire PM paradigm [111].

In this regard, one of the key aspects that needs to be tackled, in particular when considering predictive analytics – features of both PM and PPH – is the quality and heterogeneity of datasets used [112]. These data issues include amongst others: incomplete and inaccurate datasets, and limited or lack of standardized data collection systems limiting data sharing and linkage (e.g. EHRs are built differently, with limited interoperability). These altogether might result in biasing effect estimation and prediction [112].

Therefore, it is important to identify and remedy to these barriers to the collection, sharing and linkage of health-related data in order to facilitate the implementation of and expected additional health benefits that PM and PPH could offer to more traditional approaches. In this regard, there are numerous different types of barriers that influence the collection, sharing and linkage of health-related data, and these have been categorized, amongst others, in the literature as motivational, economic, legal, ethical, socio-cultural and technical barriers [113-117]. For instance, examples of key motivational and economic barriers include: (i) the risks of getting scooped of a potential publication if research data are shared with external researchers, (ii) potential damage to the reputation of the original data collectors if errors are subsequently found in the shared datasets, (iii) the belief that the data collected could be misused or misinterpreted by external researchers, (iv) the lack of adequate resources (financial, time and expertise) to support data sharing and data management activities, (v) no research incentive for resource-limited institutions to participate in data sharing activities, and (vi) academic survival [115, 118-124]. These different types of barriers to the collection, sharing and linkage of health data are investigated and discussed in depth in this thesis (chapters 3, 4 and 5).

1.5 Ethical challenges for precision health approaches

The ethical challenges posed by this developing approach to individualized care, PM, and PPH have been extensively discussed in the scientific literature [125-134]. These include, amongst others, the following ethical considerations.

Adequate and meaningful informed consent procedure

Given the increasing accessibility of new technologies (e.g. cheaper high throughput sequencing methods) allowing the consideration and integration of (multi)omics data and other data types to guide clinical practice, it has been argued that the vast amount and complexity of information that have to be provided to participants or patients is complicating enormously the informed consent procedure for PM interventions [125, 135]. Several reasons have been put

forward to try explain and delineate the challenging environment in which physicians or researchers strive to get a proper and meaningful informed consent from PM participants. These challenges include non-exhaustively: (i) the lack or limited level of genomic health literacy from study participants (e.g. the ability for PM participants to fully grasp the implications of pharmacogenetic testing), (ii) reservations regarding some findings of genomic testing (e.g. uncertain interpretation of genetic variants that limit not only an assessment of the benefits and risks of an intervention but also the decisional capacity of participants and physicians), (iii) implications for participants in case information about their condition needs to be shared with relatives if the latter could benefit from the same intervention (e.g. in the case of germline testing), (iv) the disclosure of results concerning genetic variants of unknown significance (VUS), (v) acceptance by study participants that their data are to be collected and subsequently shared for research purposes, and (vi) in the context of learning healthcare systems, the blurring of the boundaries between clinical care and research could put the patient in a delicate position of not potentially understanding whether the relevant genomic test for a particular PM intervention is being ordered to answer his or her healthcare needs, or the genomic test is serving general research purposes [125, 126, 135, 136].

Similarly, traditional informed consent procedures have also been problematic for PPH approaches, in particular for those based on the increasing availability of genomics information (e.g. screening programs [137], with consent challenges similar to PM), digital *Big Data* (where the need for consent is even debated since data are collected mostly from public domains on the internet – from Nissenbaum, as cited in [134]), and participatory approaches involving citizens who consent implicitly or explicitly to donate their health-related data (for instance by reporting symptoms of a possible disease on web-based platforms or mobile apps, using self-tracking devices, or sharing mobile phone sensor data) to guide public health practice [138]. For the latter, the traditional informed consent procedure has faced some feasibility issues and new methods of obtaining consent had to be developed. Indeed, traditional informed consent procedures are feasible – both practically and financially – only when there are limited numbers of participants for researchers or public health professionals to engage with each one of them to provide consent information and subsequently obtain their written informed consent prior to their participation [139]. However, the sheer number of online participants (potentially tens of thousands [138]) renders this traditional approach to obtaining informed consent unpractical. Therefore, electronic adaptations to the traditional informed consent procedures, known as the electronic consent or *e-consent*, had to be developed to cater for these limitations [140]. These have helped the implementation of such innovative and nascent PPH approaches. However, the

use of electronic consent procedures raises other ethically-relevant issues such as, amongst others, ensuring that participants have the necessary intellectual capacity to provide their e-consent in the first place, or confirming the identity of the online participant [141]. The informed consent issues for participatory approaches to digital public health surveillance are discussed in depth in chapter 7.

Privacy and Confidentiality

Other important ethical concerns regarding PM are those of privacy and confidentiality [125, 126, 130]. One of the many ethical challenges that PM initiatives have to dealt with is to find the right balance between two opposing but integral components to their proper functioning. Indeed, striking a balance between facilitating the sharing of comprehensive health-related data of PM participants for research purposes (for the greater good) with the need to protect their privacy and confidentiality (to ensure continued participant engagement) is a conundrum that many PM initiatives are struggling with [125, 130]. In order to streamline the whole process of data sharing while safeguarding data privacy, Blasimme et al [142] argued for the need to introduce more effective privacy-preserving technological solutions (e.g. blockchain technologies). The authors further argued that their implementation would require, "targeted public investment and the development of technical requirements for data exchange platforms to ensure regulatory compliance" [142]. This is particularly important given that although legislations have been put in place to safeguard the privacy and protect data subjects against discriminatory actions (e.g. in case of genetic discrimination), it has been noted by Wauters and Hoyweghen [143] that these legislations do not eliminate the fears that participants harbor with regard to genetic discrimination (e.g. in the context of employment or insurance, and even with the particular genetic disease itself. For instance, concerning Huntington's disease, which is often stigmatized and where no cure currently exists). Therefore, for the proper functioning of PM initiatives, it is important "that data should not be held hostage by patients' fear of discrimination" [125].

Privacy threats become also evident in PPH approaches and in the *Big Data* era, where databases are increasingly linked to one another, with the possibility of revealing comprehensive information on individuals and therefore threatening their privacy (e.g. by linking administrative records) [144]. Regarding participatory disease surveillance systems, privacy risks are multifactorial, and include amongst others: (i) the spatio-temporality of collected health-related data, and (ii) mobile phones' apps usage data [145, 146]. Privacy issues stemming from participatory disease surveillance systems are discussed in details in chapters 6 and 7.

Ethics of Inclusion, Cohort Diversity and Discrimination

An ethics of inclusion is required for PM and PPH initiatives in order to achieve one of their essential goals: achieving health equity by bridging the longstanding health inequities gap between different subgroups of the population, in particular for disadvantaged and marginalized groups (e.g. racial and ethnic minorities, and people with disabilities) [127, 147-150]. In order to achieve this important objective, underrepresented groups need to trust that they will obtain their equitable share of health benefits for their participation [127], and history has proven that this is not an easy task. Indeed, these minority groups are not only often granted suboptimal access to healthcare services, including low quality medical care [151], but they have also been plagued with marginalization, abuse and exploitation in the research domain [152]. Altogether, these have contributed to the mistrust against healthcare institutions and the research enterprise.

Indeed, well-known unethical and historical research examples are often underscored by minorities to justify their mistrust and unwillingness to participate in research activities or engage with their respective healthcare systems (e.g. the Tuskegee syphilis experiment or the case of Henrietta Lacks) [153, 154]. In addition, the underrepresentation of minorities in critical components of PM and PPH initiatives (e.g. genomic databases) is underlining the lingering threat that history will repeat itself for these groups, that is, they are going to miss the train of added health benefits brought forward by these two initiatives. For instance, an analysis of Genome Wide Association Studies (GWAS) conducted in 2016 revealed that regarding the composition of 35 million samples, the great majority were contributed by participants of European ancestry, highlighting the need to cater for the underrepresentation of minority populations [11]. Therefore, achieving cohort diversity is regarded as a fundamental objective in precision health approaches, that would allow to answer specific research questions and cater for health disparities that are meaningful and non-discriminatory to all racial, ethnic, and social strata of their respective societies. The ethics of inclusion, cohort diversity and discrimination for both PM and PPH are discussed in more details in chapters 8 and 9.

1.6 Understanding the bigger picture

The different components discussed in sections 1.1 to 1.4 are shown in Figure 1. Importantly, Figure 1 serves to provide an overview on how these components could be entwined with one another within the framework of learning healthcare systems, with deliverables from PM and PPH approaches being fed back into the system to improve clinical and public health practice. Please also note that neither does Figure 1 include all sources of *Big Data* being used in PM and PPH nor does it provide all applications of *Big Data* in these two domains. In addition, for

the sake of simplicity and relevance to learning healthcare systems, Figure 1 has also adopted the position taken by Sacristan and Dilla [42], i.e. *Small Data* are the foundations of *Big Data*.

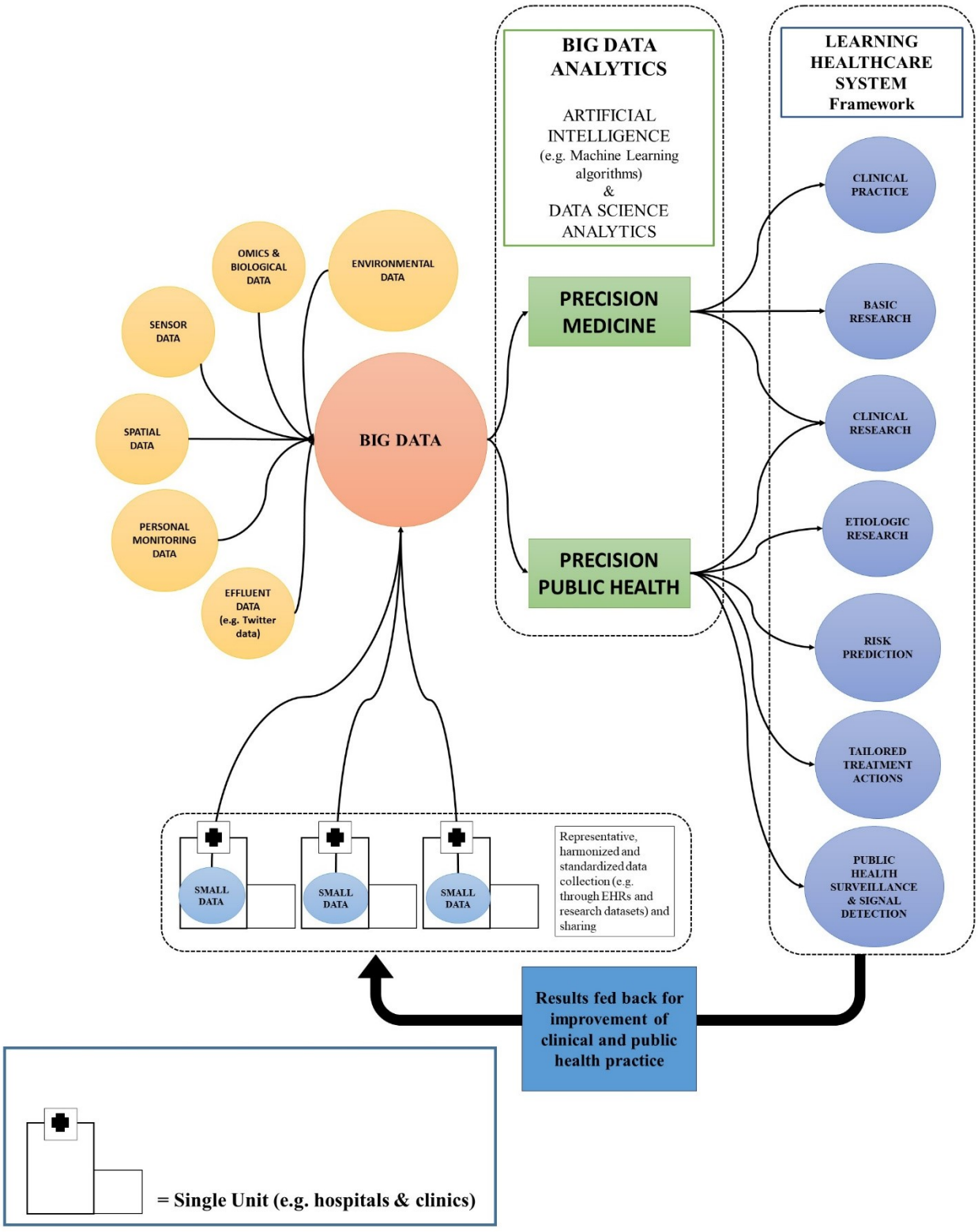


Figure 1 Relationship between components discussed in sections 1.1 to 1.4

1.7 References

1. Online Etymology Dictionary. data (n.d): Douglas Harper; 2021 [accessed 30.1.2021]. Available from: https://www.etymonline.com/word/data.
2. Cambridge Dictionary. data: Cambridge University Press; 2021 [accessed 30.1.2021]. Available from: https://dictionary.cambridge.org/fr/dictionnaire/anglais/data.
3. Murray A. CEOs: The Revolution Is Coming: Fortune; 2016 [accessed 30.1.2021]. Available from: https://fortune.com/2016/03/08/davos-new-industrial-revolution/.
4. Castells M. The information age: Oxford Blackwell Publishers; 1996.
5. Topol E. Deep Medicine. United States of America: Hachette Book Group; 2019.
6. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. Journal of Big Data. 2019;6(1):54.
7. Intersoft consulting. Recital 35 Health Data* (n.d.) [accessed 1.02.2021]. Available from: https://gdpr-info.eu/recitals/no-35/.
8. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From Big Data to Precision Medicine. Front Med. 2019;6(34).
9. Geneviève LD, Martani A, Mallet MC, Wangmo T, Elger BS. Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: A systematic review. PLoS one. 2019;14(12):e0226015.
10. Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med. 2020;26(1):29-38.
11. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature News. 2016;538(7624):161.
12. Statistica. Total amount of global healthcare data generated in 2013 and a projection for 2020: Statistica; 2018 [30.01.2021]. Available from: https://www.statista.com/statistics/1037970/global-healthcare-data-volume/.
13. Horton R. Offline: in defence of precision public health. Lancet. 2018;392(10157):1504.
14. Blumenthal D. Stimulating the Adoption of Health Information Technology. New England Journal of Medicine. 2009;360(15):1477-9.
15. Blumenthal D. Launching HITECH. N Engl J Med. 2010;362(5):382-5.
16. Schweizeriche Eidenossenschaft. Bundesgesetz über das elektronische Patientendossier: Schweizeriche Eidenossenschaft,; 2020 [accessed 30.01.2021]. Available from: https://www.fedlex.admin.ch/eli/cc/2017/203/de.

17. De Pietro C, Francetic I. E-health in Switzerland: The laborious adoption of the federal law on electronic health records (EHR) and health information exchange (HIE) networks. Health Policy. 2018;122(2):69-74.
18. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: A systematic review. J Biomed Inform. 2020;108:103500.
19. Schneider PP, van Gool CJ, Spreeuwenberg P, Hooiveld M, Donker GA, Barnett DJ, et al. Using web search queries to monitor influenza-like illness: an exploratory retrospective analysis, Netherlands, 2017/18 influenza season. Euro Surveill. 2020;25(21):1900221.
20. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of Commercially Available Wearable Activity Trackers Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods Evaluation. JMIR MHealth UHealth. 2016;4(1):e7.
21. Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. BMC Med. 2019;17(1):133-.
22. Kitchin R, McArdle G. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data Soc. 2016;3(1):2053951716631130.
23. SAS. Big Data: What it is and why it matters: SAS Institute Inc.; 2021 [accessed 4.02.2021]. Available from: https://www.sas.com/en_th/insights/big-data/what-is-big-data.html.
24. Anuradha J. A brief introduction on Big Data 5Vs characteristics and Hadoop technology. Procedia Comput Sci. 2015;48:319-24.
25. Mayer-Schonberger V, Cukier K. Big data: A revolution that will change how we live. Work and Think London: John Murray. 2013.
26. Reis-Filho JS. Next-generation sequencing. Breast Cancer Res. 2009;11(3):S12.
27. McDermott JE, Wang J, Mitchell H, Webb-Robertson B-J, Hafen R, Ramey J, et al. Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. Expert Opin Med Diagn. 2013;7(1):37-51.
28. Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. Proteomics. 2016;16(5):741-58.
29. Hall JL, Ryan JJ, Bray BE, Brown C, Lanfear D, Newby LK, et al. Merging Electronic Health Record Data and Genomics for Cardiovascular Research: A Science Advisory From the American Heart Association. Circ Cardiovasc Genet. 2016;9(2):193-202.
30. Lau-Min KS, Asher SB, Chen J, Domchek SM, Feldman M, Joffe S, et al. Real-world integration of genomic data into the electronic health record: the PennChart Genomics Initiative. Genet Med. 2020.

31. Johnson AEW, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3(1):160035.
32. Celi LA, Mark RG, Stone DJ, Montgomery RA. "Big data" in the intensive care unit. Closing the data loop. Am J Respir Crit Care Med. 2013;187(11):1157-60.
33. Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. Annu Rev Public Health. 2018;39(1):95-112.
34. Achrekar H, Gandhe A, Lazarus R, Ssu-Hsin Y, Liu B, editors. Predicting Flu Trends using Twitter data. 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS); 2011 10-15 April 2011.
35. Dai X, Bikdash M, Meyer B, editors. From social media to public health surveillance: Word embedding based clustering method for twitter classification. SoutheastCon 2017; 2017 30 March-2 April 2017.
36. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: The case of the EHR4CR project. J Biomed Inform. 2015;53:162-73.
37. King M-C, Levy-Lahad E, Lahad A. Population-Based Screening for BRCA1 and BRCA2: 2014 Lasker Award. JAMA. 2014;312(11):1091-2.
38. Domchek SM, Friebel TM, Singer CF, Evans DG, Lynch HT, Isaacs C, et al. Association of Risk-Reducing Surgery in BRCA1 or BRCA2 Mutation Carriers With Cancer Risk and Mortality. JAMA. 2010;304(9):967-75.
39. Estrin D. Small data, where n = me. Commun ACM. 2014;57(4):32–4.
40. Rabbi M, Hane Aung M, Choudhury T. Towards Health Recommendation Systems: An Approach for Providing Automated Personalized Health Feedback from Mobile Data. In: Rehg JM, Murphy SA, Kumar S, editors. Mobile Health: Sensors, Analytic Methods, and Applications. Cham: Springer International Publishing; 2017. p. 519-42.
41. Karkar R, Schroeder J, Epstein DA, Pina LR, Scofield J, Fogarty J, et al. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2017. p. 6850–63.
42. Sacristán JA, Dilla T. No big data without small data: learning health care systems begin and end with the individual patient. J Eval Clin Pract. 2015;21(6):1014-7.
43. Cofiño R, Lopez-Villar S, Suárez O. How to work with local communities to improve population health: big data and small data. J Epidemiol Community Health. 2017;71(7):657.

44. Kitchin R, Lauriault TP. Small data in the era of big data. GeoJournal. 2015;80(4):463-75.
45. Mitani AA, Haneuse S. Small Data Challenges of Studying Rare Diseases. JAMA Netw Open. 2020;3(3):e201965-e.
46. Leff DR, Yang G-Z. Big Data for Precision Medicine. Engineering. 2015;1(3):277-9.
47. Schaefer GO, Tai ES, Sun S. Precision Medicine and Big Data. Asian Bioeth Rev. 2019;11(3):275-88.
48. Dolley S. Big Data's Role in Precision Public Health. Front Public Health. 2018;6(68).
49. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health Surveill. 2020;6(2):e19509.
50. Davis JJ, Boisvert S, Brettin T, Kenyon RW, Mao C, Olson R, et al. Antimicrobial Resistance Prediction in PATRIC and RAST. Sci Rep. 2016;6:27930.
51. Omersel J, Karas Kuželički N. Vaccinomics and Adversomics in the Era of Precision Medicine: A Review Based on HBV, MMR, HPV, and COVID-19 Vaccines. J Clin Med. 2020;9(11):3561.
52. Russell SJ, Norvig P. Artificial intelligence: a modern approach. Upper Saddle River, New Jersey 07458: Pearson Education Limited; 2010.
53. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob health. 2018;3(4):e000798.
54. Statistica. Revenues from the artificial intelligence (AI) market worldwide from 2016 to 2025 (in million U.S. dollars) 2019 [accessed 2019]. Available from: https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/.
55. AIHLEG. A definition of AI: Main capabilities and scientific disciplines. European Commission; 2019.
56. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019;25(1):44-56.
57. Ross MK, Wei W, Ohno-Machado L. "Big data" and the electronic health record. Yearb Med Inform. 2014;9(1):97-104.
58. Collier M, Fu R, Yin L, Christiansen, P. ARTIFICIAL INTELLIGENCE: Healthcare's New Nervous System. 2017. Available from: https://www.accenture.com/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf

59. Thiebaut R, Thiessard F. Artificial Intelligence in Public Health and Epidemiology. Yearb Med Inform. 2018;27(1):207-10.
60. Neill DB. New Directions in Artificial Intelligence for Public Health Surveillance. IEEE Intell Syst. 2012;27(1):56-9.
61. Leelatian N, Sinnaeve J, Mistry AM, Barone SM, Brockman AA, Diggins KE, et al. Unsupervised machine learning reveals risk stratifying glioblastoma tumor cells. Elife. 2020;9:e56879.
62. Watson DS, Krutzinna J, Bruce IN, Griffiths CEM, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. BMJ. 2019;364:l886.
63. Ferretti A, Schneider M, Blasimme A. Machine Learning in Medicine. Eur Data Prot Law Rev. 2018;4(3).
64. Adibuzzaman M, DeLaurentis P, Hill J, Benneyworth BD. Big data in healthcare - the promises, challenges and opportunities from a research perspective: A case study with a model database. AMIA Annu Symp Proc. 2018;2017:384-92.
65. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. N Engl J Med. 2018;378(11):981-3.
66. Dowell SF, Blazes D, Desmond-Hellmann S. Four steps to precision public health. Nature News. 2016;540(7632):189.
67. Prosperi M, Min JS, Bian J, Modave F. Big data hurdles in precision medicine and precision public health. BMC Med Inform Decis Mak. 2018;18(1):139.
68. Group PW. The Precision Medicine Initiative Cohort Program - Building a Research Foundation for 21st Century Medicine: National Institutes of Health; 2015 [accessed 18.11.2019]. Available from: https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf.
69. Mirnezami R, Nicholson J, Darzi A. Preparing for precision medicine. N Engl J Med. 2012;366(6):489-91.
70. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372(9):793-5.
71. Ashley EA. Towards precision medicine. Nat Rev Genet. 2016;17(9):507.
72. Rubin R. A precision medicine approach to clinical trials. JAMA. 2016;316(19):1953-5.
73. Feero WG, Wicklund CA, Veenstra D. Precision Medicine, Genome Sequencing, and Improved Population Health. JAMA. 2018;319(19):1979-80.

74. Meier-Abt P, Lawrence A, Vayena E, Schwede T. The Swiss approach to precision medicine. Swiss Med Wkly. 2018.
75. PHRT. Strategic Focus Area: Personalized Health and Related Technologies (n.d.) [accessed 8.02.2021]. Available from: https://www.sfa-phrt.ch/.
76. Lawrence A, Selter L. Swiss Personalized Health Network (SPHN): L'initiative nationale en un clin d'oeil. Bull Med Suisses. 2017.
77. SDSC. SPO - Swiss Personalized Oncology (SPHN): Swiss Data Science Center; (n.d.) [accessed 1.02.2021]. Available from: https://datascience.ch/project/31777/.
78. PHRT. What is PHRT? : PHRT; (n.d.) [accessed 8.02.2021]. Available from: https://www.sfa-phrt.ch/what-is-phrt.
79. SFMPP. 1ère société savante dédiée à la médecine personnalisée: SFMPP; 2019 [accessed 8.02.2021]. Available from: https://www.sfmpp.org/presentation/.
80. AVIESAN. France Médecine Génomique 2025 2017 [accessed 8.02.2021].Available from:https://www.gouvernement.fr/sites/default/files/document/document/2016/06/22.06.2016_remise_du_rapport_dyves_levy_-_france_medecine_genomique_2025.pdf.
81. National Academies of Sciences, Engineering, Medicine. Advancing Disease Modeling in Animal-Based Research in Support of Precision Medicine: Proceedings of a Workshop. Forstag EH, Anestidou L, editors. Washington, DC: The National Academies Press; 2018. 138 p.
82. Baker DJ. Chapter 12 - Genomics England: The Future of Genomic Medical Diagnosis: Governmental Scale Clinical Sequencing and Potential Walled-Garden Impact on Global Data Sharing. In: Lambert CG, Baker DJ, Patrinos GP, editors. Human Genome Informatics: Academic Press; 2018. p. 269-92.
83. Brice P. UK Precision Medicine Catapult: what went wrong? : PHG Foundation; [accessed 8.02.2021]. Available from: https://www.phgfoundation.org/blog/UK-precision-medicine-catapult-what-went-wrong.
84. Genomics England. How we work: Department of Health and Social Care; 2020 [accessed 8.02.2021]. Available from: https://www.genomicsengland.co.uk/about-genomics-england/how-we-work/.
85. FINNGEN. FinnGen research project is an expedition to the frontier of genomics and medicine: FINNGEN; n.d. [accessed 8.02.2021]. Available from: https://www.finngen.fi/en.
86. Diatech Pharmacogenetics. Italian Society for Personalized Medicine (SIMeP): Diatech Pharmacogenetics; 2020 [accessed 8.02.2021]. Available from:

https://www.diatechpharmacogenetics.com/en/societa-italiana-medicina-personalizzata-simep/.
87. Metspalu A. The Estonian Genome Project. Drug Dev Res. 2004;62(2):97-101.
88. NIH. The future of health begins with you. : National Institutes of Health: All of Us Research Progra; 2021 [accessed 8.02.2021]. Available from: https://allofus.nih.gov/.
89. Columbia University Institute for Genomic Medicine. Precision Medicine Initiatives: Columbia University; [accessed 8.02.2021]. Available from: https://www.igm.columbia.edu/precision-medicine-initiatives.
90. Hoyle B. Giga-speed bioinformatics to power Genome Canada. Nat Biotechnol. 1999;17(10):950-.
91. Genome Canada. Collaborating on the future Annual Report 2019-20. Ottawa: Genome Canada; (n.d.). Available from: https://www.genomecanada.ca/sites/default/files/genomecanada-annualreport-en-web.pdf
92. Ontario Genomics. Ontario Personalized Medicine Network (OPMN): Ontario Genomics; 2020 [accessed 8.02.2021]. Available from: https://www.ontariogenomics.ca/provincial-strategies/opmn/.
93. Génome Québec. Personalized Medicine: Génome Québec Inc; 2021 [cited 2021 8.02.2021]. Available from: https://www.genomequebec.com/en/health/personalized-medicine/.
94. Bajaj A. The Genome India Project: Invest India; 2020 [accessed 8.02.2021]. Available from: https://www.investindia.gov.in/team-india-blogs/genome-india-project.
95. JapanGov. Optimal Medical Treatment for Everyone: The Government of Japan; n.d. [accessed 9.02.2021]. Available from: https://www.japan.go.jp/tomodachi/2018/Autumn2018/optimal_medical_treatment_for_everyone.html.
96. Lasalvia L, Merges R. Expanding Precision Medicine. Journal of Precision Medicine. 2019;5:1-5.
97. H3Africa. H3Africa: Human Heredity and health in Africa 2021 [accessed 9.02.2021]. Available from: https://h3africa.org/.
98. Khoury MJ, Galea S. Will Precision Medicine Improve Population Health? JAMA. 2016;316(13):1357-8.
99. Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. BMC Med Ethics. 2020;21(1):17.

100. Weeramanthri TS, Dawkins HJS, Baynam G, Bellgard M, Gudes O, Semmens JB. Editorial: Precision Public Health. Front Public Health. 2018;6(121).
101. Khoury MJ, Iademarco MF, Riley WT. Precision Public Health for the Era of Precision Medicine. Am J Prev Med. 2016;50(3):398-401.
102. Khoury MJ, Armstrong GL, Bunnell RE, Cyril J, Iademarco MF. The intersection of genomics and big data with public health: Opportunities for precision public health. PLoS Med. 2020;17(10):e1003373.
103. World Health Organization. Public health surveillance 2021 [accessed 9.02.2021]. Available from: https://www.who.int/immunization/monitoring_surveillance/burden/vpd/en/.
104. Khoury MJ, Engelgau M, Chambers DA, Mensah GA. Beyond Public Health Genomics: Can Big Data and Predictive Analytics Deliver Precision Public Health? Public Health Genomics. 2018;21(5-6):244-50.
105. Armstrong GL, MacCannell DR, Taylor J, Carleton HA, Neuhaus EB, Bradbury RS, et al. Pathogen Genomics in Public Health. N Engl J Med. 2019;381(26):2569-80.
106. Thiel M, Parkin J, Dratwa J, Halila R, Palazzani K, A G. The ethical implications of new health technologies and citizen participation Luxembour: Publications office of the European Union; 2016 [accessed 1.10.2017]. Available from: https://publications.europa.eu/en/publication-detail/-/publication/e86c21fa-ef2f-11e5-8529-01aa75ed71a1.
107. Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the Importance of Digital Epidemiology. N Engl J Med. 2013;369(5):401-4.
108. Smolinski MS, Crawley AW, Olsen JM, Jayaraman T, Libel M. Participatory Disease Surveillance: Engaging Communities Directly in Reporting, Monitoring, and Responding to Health Threats. JMIR Public Health Surveill. 2017;3(4):e62.
109. Chan EH, Brewer TF, Madoff LC, Pollack MP, Sonricker AL, Keller M, et al. Global capacity for emerging infectious disease detection. Proceedings of the National Academy of Sciences. 2010;107(50):21701.
110. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big Data for Infectious Disease Surveillance and Modeling. J Infect Dis. 2016;214(suppl_4):S375-S9.
111. Whitcomb DC. Barriers and Research Priorities for Implementing Precision Medicine. Pancreas. 2019;48(10):1246-9.
112. Shah ND, Steyerberg EW, Kent DM. Big Data and Predictive Analytics: Recalibrating Expectations. JAMA. 2018;320(1):27-8.

113. Mostert M, Bredenoord AL, Biesaart MC, Van Delden JJ. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet. 2016;24(7):956-60.
114. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, et al. Making sense of big data in health research: towards an EU action plan. Genome Med. 2016;8(1):1-13.
115. Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC public health. 2014;14(1):1-9.
116. Edwards A, Hollin I, Barry J, Kachnowski S. Barriers to cross--institutional health information exchange: a literature review. Journal of healthcare information management: JHIM. 2010;24(3):22-34.
117. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc. 2010;17(3):288-94.
118. Carr D, Littler K. Sharing research data to improve public health: A funder perspective. J Empir Res Hum Res Ethics. 2015;10(3):314-6.
119. Mladovsky P, Mossialos E, McKee M. Improving access to research data in Europe. BMJ; 2008;336(7639):287-8
120. Popkin G. Data sharing and how it can benefit your scientific career. Nature; 2019;569(7756):445-447
121. Bierer BE, Crosas M, Pierce HH. Data Authorship as an Incentive to Data Sharing. N Engl J Med. 2017;376(17):1684-7.
122. Fecher B, Friesike S, Hebing M. What Drives Academic Data Sharing? PLoS one. 2015;10(2):e0118053.
123. Institute of Medicine, Board on Health Sciences, Board on Health Sciences Policy, Roundtable on Translating Genomic-Based Research for Health, National Cancer Policy Forum, Forum on Neuroscience and Nervous System Disorders, et al. Sharing Clinical Research Data: Workshop Summary. Downey AS, Olson S, editors: National Academies Press; 2013. 156 p.
124. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PloS one. 2015;10(8):e0134826-e.
125. Fiore RN, Goodman KW. Precision medicine ethics: selected issues and developments in next-generation sequencing, clinical oncology, and ethics. Curr Opin Oncol. 2016;28(1).

126. Minari J, Brothers KB, Morrison M. Tensions in ethics and policy created by National Precision Medicine Programs. Human Genomics. 2018;12(1):22.
127. Lee SS-J, Fullerton SM, Saperstein A, Shim JK. Ethics of inclusion: Cultivate trust in precision medicine. Science. 2019;364(6444):941.
128. Hammer MJ, editor Precision medicine and the changing landscape of research ethics. Oncology nursing forum; 2016: Oncology Nursing Society.
129. Blasimme A, Vayena E. Becoming partners, retaining autonomy: ethical considerations on the development of precision medicine. BMC Med Ethics. 2016;17(1):1-8.
130. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. Genet Med. 2017;19(7):743-50.
131. Williams JS, Walker RJ, Egede LE. Achieving Equity in an Evolving Healthcare System: Opportunities and Challenges. Am J Med Sci. 2016;351(1):33-43.
132. Juengst ET, Van Rie A. Transparency, trust, and community welfare: towards a precision public health ethics framework for the genomics era. Genome Med. 2020;12(1):98.
133. Susser D. Ethical Considerations for Digitally Targeted Public Health Interventions. Am J Public Health. 2020;110(S3):S290-S1.
134. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical Challenges of Big Data in Public Health. PLoS Comp Biol. 2015;11(2):e1003904.
135. McGowan ML, Settersten RA, Jr., Juengst ET, Fishman JR. Integrating genomics into clinical oncology: ethical and social challenges from proponents of personalized medicine. Urol Oncol. 2014;32(2):187-92.
136. Hunter DJ. Uncertainty in the Era of Precision Medicine. N Engl J Med. 2016;375(8):711-3.
137. Goldenberg AJ, Sharp RR. The Ethical Hazards and Programmatic Challenges of Genomic Newborn Screening. JAMA. 2012;307(5):461-2.
138. Geneviève LD, Martani A, Wangmo T, Paolotti D, Koppeschaar C, Kjelsø C, et al. Participatory Disease Surveillance Systems: Ethical Framework. J Med Internet Res. 2019;21(5):e12273.
139. Elger BS, Caplan AL. Consent and anonymization in research involving biobanks: differing terms and norms present serious barriers to an international framework. EMBO Rep. 2006;7(7):661-6.
140. Coiera E, Clarke R. e-Consent: The design and implementation of consumer consent mechanisms in an electronic environment. J Am Med Inform Assoc. 2004;11(2):129-40.

141. Grady C, Cummings SR, Rowbotham MC, McConnell MV, Ashley EA, Kang G. Informed Consent. N Engl J Med. 2017;376(9):856-67.
142. Blasimme A, Fadda M, Schneider M, Vayena E. Data Sharing For Precision Medicine: Policy Lessons And Future Directions. Health Affairs. 2018;37(5):702-9.
143. Wauters A, Van Hoyweghen I. Global trends on fears and concerns of genetic discrimination: a systematic literature review. J Hum Genet. 2016;61(4):275-82.
144. Brown AP, Ferrante AM, Randall SM, Boyd JH, Semmens JB. Ensuring Privacy When Integrating Patient-Based Datasets: New Methods and Developments in Record Linkage. Front Public Health. 2017;5(34).
145. Olvingson C, Hallberg J, Timpka T, Lindqvist K. Ethical issues in public health informatics: implications for system design when sharing geographic information. J Biomed Inform. 2002;35(3):178-85.
146. Welke P, Andone I, Blaszkiewicz K, Markowetz A. Differentiating smartphone users by app usage. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Heidelberg, Germany: Association for Computing Machinery; 2016. p. 519–23.
147. Hindorff LA, Bonham VL, Ohno-Machado L. Enhancing diversity to reduce health information disparities and build an evidence base for genomic medicine. Per Med. 2018;15(5):403-12.
148. Sabatello M. Precision medicine, health disparities, and ethics: the case for disability inclusion. Genet Med. 2018;20(4):397-9.
149. Kee F, Taylor-Robinson D. Scientific challenges for precision public health. J Epidemiol Community Health. 2020;74(4):311.
150. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring Fairness in Machine Learning to Advance Health Equity. Ann Intern Med. 2018;169(12):866-72.
151. Bhopal RS. Racism in health and health care in Europe: reality or mirage? Eur J Public Health. 2007;17(3):238-41.
152. Cohn EG, Henderson GE, Appelbaum PS. Distributive justice, diversity, and inclusion in precision medicine: what will success look like? Genet Med. 2017;19(2):157-9.
153. Lee SS-J, Cho MK, Kraft SA, Varsava N, Gillespie K, Ormond KE, et al. “I don’t want to be Henrietta Lacks”: Diverse patient perspectives on donating biospecimens for precision medicine research. Genet Med. 2019;21(1):107-13.
154. Gamble VN. Under the shadow of Tuskegee: African Americans and health care. Am J Public Health. 1997;87(11):1773-8.

Chapter 2: Research Objectives and Methodology

“Art and science have their meeting point in method” - Edward G. Bulwer-Lytton

2.1 Research objectives

This thesis forms part of a larger research project, launched within the framework of the National Research Programme 74 (NRP74) on *Smarter Health Care*, titled "Advancing SMart solutions and eliminating barriers for the Acquisition, Analysis, and Sharing of Health data in Switzerland" (SMAASH). The SMAASH project was funded by the Swiss National Science Foundation. The main objectives of this project were, amongst others, to identify barriers and facilitators (e.g. ethical, legal, political, etc.) that influenced the processing of health data in the Swiss context, and to provide insights on Swiss stakeholders' attitudes and expertise regarding these factors. Therefore, this thesis was conceived partly within these boundaries, and it aims to explore issues arising from the processing of data for PM and PPH, two domains particularly important for the evolving Swiss healthcare and public health landscapes. The thesis also examines issues of fairness, discrimination and justice that could arise from the collection, sharing and use of non-harmonized and low quality datasets in PM and PPH.

This thesis addresses the following research questions:

1. **What are the barriers and facilitators to the processing of health data in Switzerland?**

   Health data are the foundations of precision health approaches. For PM and PPH to achieve their goals of promoting and safeguarding individual and population health, it is crucial that the collected health data are not only of high quality but also easily shareable between different stakeholders involved in shaping the health of individuals. In this perspective, it is paramount to identify barriers and facilitators that influence the processing of health data, in particular data sharing, so that prompt policy-making solutions can be implemented. This important question for health policy is addressed in chapters 3, 4 and 5 of this thesis.

   Chapter 3 is concerned with the systematic identification and analysis of existing literature on barriers and facilitators to the processing of health data encountered by Danish and Swiss multi-center studies. The multi-center criterion was important to identify practices that contribute or hinder harmonization approaches to the processing of health data. In addition, Denmark was chosen as a comparison country because of its long tradition of nationwide registries and the relative ease that Danish researchers carry out data sharing and linkage activities. Therefore, success mechanisms in Denmark were assessed for their potential implementation in the Swiss context.

Chapter 4 investigated, through semi-structured interviews, the fairness of the Swiss academic and legal systems towards expert stakeholders (mostly researchers) in their data sharing activities, providing another perspective on the fairness dimension of data sharing where the focus was generally up-to-now on data subjects or on the reproducibility of research findings. This chapter is extremely relevant to guide policy-making decisions, in particular following the impetus given to the open science and data sharing movements for research carried out in Switzerland. Indeed, it explores the systemic contradictions and legal uncertainty that hinder the sharing of health data between Swiss expert stakeholders, with the ultimate goal of informing policy-making and promoting the sharing of health data in Switzerland.

Putting aside the open science involving data repositories and systemic constraints to data sharing activities, chapter 5 explores and assesses the fairness of the interpersonal negotiation process between the original data collectors and data recipients for sharing health-related data (i.e. through controlled data access). It has gathered the perspectives of expert stakeholders through semi-structured interviews and identified via a distributive justice lens, namely desert-based principles, data sharing practices that could be perceived as fair or unfair by the parties involved, and therefore those practices that are likely to foster data sharing. These practices were then assessed ethically and recommendations were made accordingly.

2. **What are the ethical, legal and social issues emerging from participatory disease surveillance systems and how are research ethics coping with this rapidly evolving precision public health field?**

   The novel coronavirus pandemic, COVID-19, has shown that in our increasingly connected world, it is easier for epidemics to cross national borders whilst inflicting significant damages to the health of populations and the global economy. In this perspective, it is important that data that can be useful in guiding and improving the efficiency of public health practice – both nationally and internationally – are not hindered unnecessarily by ethical and legal considerations that negatively impact the harmonized collection and sharing of health data for PPH approaches. Using the Influenzanet Consortium as a case study – a well-established European participatory disease surveillance system for Influenza-like illness (ILI) – this thesis investigates how research ethics are being implemented by these innovative disease surveillance systems

that are complementary to more traditional approaches, and thereafter, suggests realistic means of improving their ethical approaches in chapters 6 and 7.

Chapter 6 investigates, through a systematic search and analysis of existing literature on Influenzanet, how research ethics are being handled by the different country-specific ILI surveillance systems of the consortium. The overall goal is to identify areas where ethical approaches could be harmonized and improved between the different country-specific platforms. Indeed, the identified ethical, legal and social issues (ELSIs) from the literature search were then analyzed, and solutions to strengthen and maximize the positive impact of these participatory disease surveillance platforms were formulated.

Chapter 7 provides a more in-depth theoretical assessment of the ELSIs identified in chapter 6, discussing issues of informed consent and its adaptation to the digital world, protection of participants' privacy, ensuring justice and lastly, the need for capacity-building of research ethics committees to act as safety nets for digital public health surveillance. Chapter 7 ends with an ethical framework to guide public health practice for participatory disease surveillance systems.

3. **How are precision health approaches vulnerable to structural racism and what solutions could be implemented to ensure health equity and social justice?**
   This thesis has also investigated theoretically, in chapters 8 and 9, how PPH and PM could be particularly vulnerable to the more insidious impact of structural racism in healthcare, public health and research domains. Chapters 8 and 9 are important to better comprehend how PPH and PM could adhere to their health equity and social justice goals whilst preventing the widening of the health inequities gap between racial and ethnic groups.

   Taking the COVID-19 pandemic as an example, chapter 8 discusses not only how structural racism has been contributing to the disproportionate impact of the novel coronavirus pandemic on ethnic and racial minorities but also how PPH approaches might be biased towards minority groups through bias encoded in datasets or the data crisis on racial and ethnic minorities undermining public health interventions. Furthermore, the resulting data racism and its influences on machine learning algorithms guiding PPH approaches are discussed.

Similarly, chapter 9 aims to forecast the impact of structural racism in PM initiatives, and it is among the first publications to do so in the academic literature. Indeed, chapter 9 discusses three nodes of a process flow where structural racism could influence and biased the collection, integration of health data from ethnic and racial minorities in PM initiatives, and their deliverables. Chapter 9 also provides a series of recommendations and potential actions that could help reduce the impact of structural racism on PM initiatives.

2.2 Methodology

The objectives of this thesis were achieved through both empirical and theoretical research methods (Figure 2).

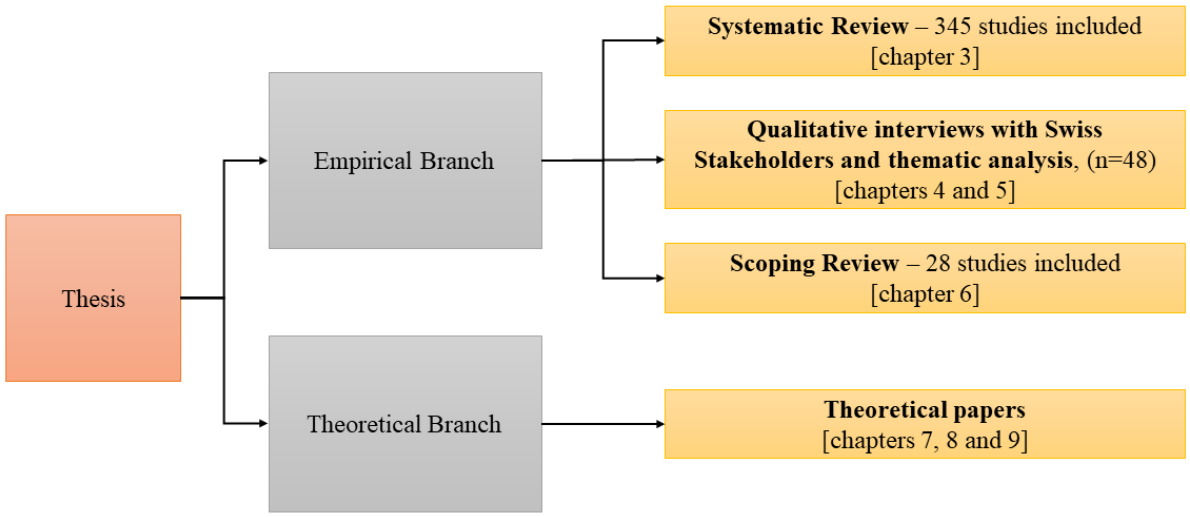


Figure 2 Empirical and theoretical branches of thesis

Chapter 3

Search strategy and study selection: The objectives of chapter 3 were fulfilled through a systematic review that adheres to the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines [1]. Its protocol was registered on PROSPERO, an international register for systematic reviews [2], in 2018 (PROSPERO ID: CRD42018081424). A systematic search strategy was developed, adapted and performed on PubMed, CINAHL, Embase (excluding Medline) and Web of Science (all databases) to retrieve publications with publication dates ranging from January 2008 to March 2019 (the initial search was initially done for publication dates ranging from January 2008 to December 2017, but it had to be repeated to include more recent publications following the request made by the journal editor). Additionally, the reference lists of included publications were screened to identify further

publications that fulfilled the inclusion criteria for the review. An example of a search strategy developed for identifying Swiss projects on PubMed was:

> ("Administrative Claims, Healthcare"[Mesh] OR "Health Records, Personal"[Mesh] OR "Clinical Coding"[Mesh] OR "Patient Discharge Summaries"[Mesh] OR "Clinical Trials as Topic"[Mesh]) AND ("Databases as Topic"[Mesh] OR "Data Collection"[Mesh] OR "Medical Informatics"[Mesh] OR "Medical Record Linkage"[Mesh] OR "Information Dissemination"[Mesh] OR "Data Integration" OR "Data Sharing") AND ("Switzerland"[Mesh]).

It can be noted that the concept of harmonization was not included in the search strategy due to its ill-defined limits [3] and after noticing that its inclusion as an essential component had severely reduced the number of publications per country (i.e. for Denmark and Switzerland).

The eligibility criteria for included studies for this review were:

(i) Publications dealing with studies processing health data (e.g. collecting, sharing or linking health datasets) were included. There were no restrictions imposed on the design or type of studies (e.g. studies with quantitative, qualitative or mixed methodologies were included). However, systematic reviews were excluded;

(ii) The data processing activities should occur between institutions, or at cross-regional or cross-national levels with the involvement of at least collaborators from either Denmark or Switzerland;

(iii) There were no limitations or restrictions based on study participants' ethnic background, diseases, gender or age.

(iv) Publication date ranging from January 2008 to March 2019.

Data extraction and quality assessment: Data extraction – in the form of title-and-abstract, full-text and reference screenings – was carried out by three review authors (Andrea Martani, Maria Christina Mallet and myself) using a standardized data extraction form developed for this study. Tenzin Wangmo, another review author, validated randomly one fifth of the publications from which the three previous authors extracted data for quality assessment purposes. We deemed acceptable to have a disagreement level of less than 10% for data entries.

Analysis: For this review, a narrative synthesis was carried out [4], and involved the grouping of the identified projects on the following grounds: source from which projects collected health data, their national or international scope, and the identified factors influencing the processing

of health data, i.e. the barriers and facilitators. The analysis was carried out using STATA ® (version 15.0) [5].

More extensive details regarding the methodology are available in chapter 3.

Chapters 4 and 5

Ethics statement: No ethics approval was necessary for these studies, since their data was collected as part of a bigger research project at the Swiss National Research Programme 74[12] on Smarter Health Care, titled "advancing SMart solutions and eliminating barriers for the Acquisition, Analysis, and Sharing of Health data in Switzerland" (SMAASH). The SMAASH project is not within the scope of the Swiss Human Research Act[13], which was confirmed by the cantonal ethics committee in Northwest and Central Switzerland (EKNZ). The ethics application number for the SMAASH project is: EKNZ req-2017-00810.

Interview guide: An interview guide was developed by Bernice Simone Elger, Tenzin Wangmo and myself. It was then pilot tested to ensure that questions were easily understandable and not ambiguous.

Participant identification and recruitment: Eligibility criteria for participation in this study included that study participants were active in the Swiss healthcare and research domains, either as researchers, policy-makers or as individuals with senior positions, who are managing health data enterprises (e.g. registries or initiatives around improving the quality and processing of health datasets). They were predominantly recruited via purposive sampling [6] (some were also identified through studies gathered from the systematic review [7] carried out in chapter 3) but also through snowball sampling [8], where some interviewees were asked to recommend other potential interviewees. The participants were recruited via email and they provided their oral consent prior to being interviewed.

Data collection and analysis: Semi-structured interviews were then conducted with the expert stakeholders on an agreed date, either in person or phone/skype, depending on their availability. Interviews were audio-recorded, transcribed verbatim with personally-identifiable information removed to ensure their confidentiality. Thematic analysis [9] was then conducted using the qualitative data analysis software, MAXQDA [10], which helped in the development of a

[12] Swiss National Science Foundation (2021): "Smarter Health Care National Research Programme". URL: http://www.nfp74.ch/en/Pages/Home.aspx. [Date accessed: 15.02.2021]
[13] The Federal Council (2020): "Federal Act on Research involving Human Beings". URL: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html. [Accessed date: June 5, 2020]

coding tree and the identification of themes and subthemes pertinent to the research objectives of chapters 4 and 5.

More extensive details regarding the methodology are available in chapters 4 and 5.

Chapter 6

Search strategy and study selection: The guidance for conducting systematic scoping reviews provided by Peters and colleagues [11] was followed for this chapter. PubMed, Global Digital Library on Ethics, Web of Science (all databases) and BELIT are the four databases that were searched for this study. The names of the country-specific platforms of the Influenzanet Consortium were used as search terms to retrieve publications, with publication dates ranging from 2003 to 2017. To be eligible for this study, publications needed to include at least one of the six ELSI[14]s considered for this study, namely (i) study approval by a research ethics committee (REC), (ii) participation should be non-discriminative and open to the general population, (iii) an information sheet should be provided on the country-specific platforms, (iv) the presence of e-consent procedures, (v) the possibility for study participants to opt out, and lastly, (vi) the presence of data security measures to protect personal data.

Data extraction, quality assessment and analysis: The methodological process for selecting publications in this study also adheres to the PRISMA guidelines [1]. Title-and-abstract, full-text and reference screenings were carried out independently by Tenzin Wangmo and I, and discrepancies between data entries or analysis were solved between us. Moreover, when clarifications or information were missing on certain platforms, their respective representatives were contacted for additional inputs.

More extensive details on the methodology are provided in chapter 6.

Chapters 7, 8 and 9

Methodological approach**:** A review of the literature adapted to the context of each research question or phenomenon described in chapters 7, 8 and 9 was conducted to identify key publications that would help in analyzing the problem at stake, and all of its potential ramifications. In contrast to chapters 8 and 9, chapter 7 used the participatory disease surveillance system, Influenzanet, as a case study. This was made possible thanks to trustful collaborations that the authors established between the Institute for Biomedical Ethics at the University of Basel and members of the Influenzanet Consortium.

[14] Acronym for: Ethical, Legal and Social issues

2.3 Author Contributions

The author contributions for chapters 3 to 9 are detailed in their respective chapters. Prof. Bernice Elger conceived and supervised with the help of Dr. Tenzin Wangmo the different research steps of the SMAASH project. Bernice Elger received the necessary funding for the latter project from the Swiss National Science Foundation (grant number: 407440_167356). I was not involved in the project application for ethical approval at the EKNZ. The interview guide was developed by Tenzin Wangmo, Bernice Elger and myself.

2.4 References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS med. 2009;6(7):e1000097.
2. National Institute for Health Research (NIHR). PROSPERO: International prospective register of systematic reviews: Centre for Reviews and Dissemination, University of York; 2018 [accessed 2.02.2021]. Available from: https://www.crd.york.ac.uk/prospero/.
3. Schmidt B-M, Colvin CJ, Hohlfeld A, Leon N. Defining and conceptualising data harmonisation: a scoping review protocol. Syst Rev. 2018;7(1):1-6.
4. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version. 2006;1:b92.
5. STATA. Your data tell a story: StataCorp LLC; 2021 [3.02.2021]. Available from: https://www.stata.com/.
6. Onwuegbuzie AJ, Leech NL. Sampling designs in qualitative research: Making the sampling process more public. Qual Rep. 2007;12(2):238-54.
7. Geneviève LD, Martani A, Mallet MC, Wangmo T, Elger BS. Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: A systematic review. PloS one. 2019;14(12):e0226015.
8. Naderifar M, Goli H, Ghaljaie F. Snowball sampling: A purposeful method of sampling in qualitative research. Stride Dev Med Educ. 2017;14(3).
9. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77-101.
10. MAXQDA. The Best Choice for Your Qualitative & Mixed Methods Research: MAXQDA; 2021 [3.02.2021]. Available from: https://www.maxqda.com/.
11. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc. 2015;13(3):141-6.

Chapter 3: Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: A systematic review

Lester Darryl Geneviève[1], Andrea Martani[1], Maria Christina Mallet[2], Tenzin Wangmo[1], Bernice Simone Elger[1,3]

[1] Institute for Biomedical Ethics, University of Basel, Basel, Switzerland,

[2] Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland,

[3] University Center of Legal Medicine, University of Geneva, Geneva, Switzerland

Citation: Geneviève LD, Martani A, Mallet MC, Wangmo T, Elger BS (2019) Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: A systematic review. PLoS ONE 14(12): e0226015. DOI: 10.1371/journal.pone.0226015 [CC-BY 4.0]

"As medical data has such power to deliver better understanding of disease and better patient outcomes, it is important we find the best way of sharing it." – Mark Walport

3.1 Abstract

Introduction

The digitalization of medicine has led to a considerable growth of heterogeneous health datasets, which could improve healthcare research if integrated into the clinical life cycle. This process requires, amongst other things, the harmonization of these datasets, which is a prerequisite to improve their quality, re-usability and interoperability. However, there is a wide range of factors that either hinder or favor the harmonized collection, sharing and linkage of health data.

Objective

This systematic review aims to identify barriers and facilitators to health data harmonization—including data sharing and linkage—by a comparative analysis of studies from Denmark and Switzerland.

Methods

Publications from PubMed, Web of Science, EMBASE and CINAHL involving cross-institutional or cross-border collection, sharing or linkage of health data from Denmark or Switzerland were searched to identify the reported barriers and facilitators to data harmonization.

Results

Of the 345 projects included, 240 were single-country and 105 were multinational studies. Regarding national projects, a Swiss study reported on average more barriers and facilitators than a Danish study. Barriers and facilitators of a technical nature were most frequently reported.

Conclusion

This systematic review gathered evidence from Denmark and Switzerland on barriers and facilitators concerning data harmonization, sharing and linkage. Barriers and facilitators were strictly interrelated with the national context where projects were carried out. Structural changes, such as legislation implemented at the national level, were mirrored in the projects. This underlines the impact of national strategies in the field of health data. Our findings also suggest that more openness and clarity in the reporting of both barriers and facilitators to data harmonization constitute a key element to promote the successful management of new projects

using health data and the implementation of proper policies in this field. Our study findings are thus meaningful beyond these two countries.

3.2 Introduction

Technological advances made over the past few years have increased the digitalization of medicine, thus leading to a considerable growth of clinical, research and public health datasets. These data sources are increasingly related to the big data environment and they include, amongst others, genomics and other-omics related-data collections, electronic health records (EHRs), patient registries, medical imaging, administrative data and clinical trials data [1, 2]. However, a good part of such datasets are often kept and analysed in silos and not adequately shared [3]. If properly integrated into the clinical life cycle, such collections of data stand to offer a unique opportunity to drive scientific discoveries and improve healthcare research. For example, they may allow a better understanding of the aetiology of illnesses and subsequently help in improving the management, prevention and treatment of diseases [1, 2]. This is even more promising in the framework of learning healthcare systems, where clear boundaries between research and care are dissolving and the same data are used both for improving scientific knowledge and providing better care [4].

In this context, developing the harmonization of health data—described as the sum of all "efforts to combine data from different sources and provide users with a comparable view of data from different studies" [5]—is crucial to improve clinical research and practice. Such standardized approach requires not only better quality, re-usability and interoperability of data, but also more open and collaborative communication between the different stakeholders active in the health data environment [6]. The fact that a good percentage of healthcare spending are being wasted as a consequence of under-exploiting data potential in several healthcare systems around the world [7–9] should be considered as one important factor urging for such changes to happen. Harmonized health datasets are also laying the foundation of a new era of biomedical research, where three concepts are currently converging, namely precision medicine, learning healthcare systems and implementation science [7, 10].

The harmonization of health data is a complex procedure which involves significant changes in how data are collected, shared and linked. Harmonization can be either prospective, when modifications occur in the study design to subsequently render the pooling of data more straightforward, or retrospective, when pooling is performed with data collected previously according to different study designs [11]. In practical terms, harmonization can be achieved through two distinct but complementary approaches, namely a "stringent" and a "flexible one"

[12]. By means of a "stringent" approach, data are harmonized through the use of standard collection tools and standard operating procedures, implementable only in a prospective way. With the "flexible" approach, on the contrary, different data collection tools might be used, as long as operating procedures are standardized [12].

In achieving the harmonization of health data, careful consideration needs to be given to already well-known as well as novel challenges related to the processes of collection, sharing and linkage. Such challenges are drastically intensified by the vastness and the hyper-connectedness of data at present time [13], which may result in unforeseen connections or crossreferencing between datasets, drastically increasing re-identification risks for data subjects [14]. The presence of these challenges has resulted in the emerging of several barriers to the effective use and sharing of health-related data [2, 15]. Although these have been categorized as technical, motivational, economic, political, socio-cultural, ethical and legal [1, 2, 15–17], a more precise mapping of the exact content of such barriers, and of the solutions that have been elaborated to mitigate them, is lacking.

Within this framework, the aim of this systematic review is to identify more precisely some of the barriers and facilitators encountered in the effort to achieve harmonization of health data—including the processes of data sharing and linkage—by a comparative analysis of studies conducted in two countries having different healthcare systems and data infrastructures, namely Denmark and Switzerland. These countries where chosen because, although they both offer high quality healthcare, they have two very diverse healthcare systems and two different data infrastructure models for healthcare. Denmark has a Beveridge-based national healthcare system [18] and a long tradition of data linkage in health through its nationwide registries [19]. On the contrary, Switzerland is based on a federalist Bismarckian organization of healthcare [20, 21] and started much later to develop strategies in the field of Health Information Exchange [22]. In this perspective, this review seeks to identify past and current studies related to the field of harmonized health data collection, sharing and linkage in these two countries and list the barriers encountered and the facilitators that make these projects successful. Furthermore, the review aims to provide some insights on the complexities associated with the use of health data that can be of relevance also in the broader international context.

3.3 Methodology

Search strategy and study selection

This study conformed to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [23], and its protocol was registered on January 3rd 2018 on

PROSPERO (CRD42018081424). A systematic literature search was performed on four search engines and electronic bibliographic databases namely PubMed, Web of Science (all databases), EMBASE (no Medline) and CINAHL for publications with dates ranging from 1st January 2008 to 31st December 2017. The time period is aligned with the adoption of the Swiss national eHealth Strategy in 2007, with the aim of introducing electronic patient records at national level [24]. The search was repeated for the period of 1st January 2018 to 31st March 2019 to include additional publications and to ensure that our systematic review is up-to-date. Reference lists of included publications were screened to identify other potential harmonized health data collection, sharing or linkage projects. A search strategy was developed for each electronic database. The literature search included Medical Subject Headings (MeSH) terms and free applicable text to health data collection, sharing and linkage. The literature search included Medical Subject Headings (MeSH) terms and free applicable text to health data collection, sharing and linkage. The search strategy consisted of three components, namely (1) types of health data, (2) keywords for data collection, sharing and linkage and (3) country of interest. For instance, the search strategy for Switzerland on PubMed was: ("Administrative Claims, Healthcare"[Mesh] OR "Health Records, Personal"[Mesh] OR "Clinical Coding"[Mesh] OR "Patient Discharge Summaries"[Mesh] OR "Clinical Trials as Topic"[Mesh]) AND ("Databases as Topic"[Mesh] OR "Data Collection"[Mesh] OR "Medical Informatics"[Mesh] OR "Medical Record Linkage"[Mesh] OR "Information Dissemination"[Mesh] OR "Data Integration" OR "Data Sharing") AND ("Switzerland"[Mesh]) [filters used are Articles types (Clinical Study, Clinical Trials (including controlled and Phases I to IV), Comparative Study, Evaluation Studies, Journal Article, Multicenter Study, Observational Study, Pragmatic Clinical Trial, Randomized Control Trial, Technical Report and Validation Studies), language (Danish, English, French and German) and species (Human Studies)]. We did not include harmonization as an imperative component in our search strategy since the exact boundaries of this concept are still controversial [25] and the addition of the term “harmonization” as an imperative component drastically reduced the number of publications for each country.

Eligibility criteria for this study were: (i) publications based on health data collection, sharing or linkage projects. There was no restriction on study design and type, i.e. qualitative, quantitative or mixed method studies, and clinical or observational studies were included; systematic reviews were excluded; (ii) there were no restriction on age, gender, disease and ethnic group of participants involved in these studies; (iii) the studies had to involve some health data collection, sharing or linkage at cross-institutional, cross-national or cross-regional levels

in at least one of the two countries; (iv) only English, French, German and Danish language articles were included, and (v) publication year of articles ranged from January 2008 to March 2019.

Data extraction and quality assessment

The literature search results were catalogued on EndNote™ X8, a reference manager software. The titles and abstracts of all articles were screened independently by two authors (LDG and AM). The full-texts of the included publications were reviewed by LDG and AM to ensure that they met the eligibility criteria to be included in the systematic review. LDG and AM performed independently the data extraction from the included articles through a standard data extraction form developed progressively by the authors of this review. Additional publications gathered through reference screening went through title and abstract, and independent full-text screenings and data extraction by MCM. Another review author, TW, validated randomly twenty percent of the publications reviewed by LDG, AM and MCM, to assess the quality of the data extraction process. A disagreement level of less than 10% for the data entries was considered acceptable.

The data extraction form included (i) study information (author(s) and publication year), (ii) sources of health data, (iii) cross-institutional or cross-national nature of the study, (iv) presence or absence of primary and secondary health data collection, analysis and sharing, and lastly (v) the categorization of barriers and facilitators to harmonized health data collection, sharing and linkage. The sources of health data were categorized as having three standard origins, namely health services, public health and research [26]. Other sources of health data falling outside these three categories were classified in a residual category ("Other").

LDG and AM performed a categorization of the identified barriers and facilitators separately, and came to consensus on the final categorization of these elements for accuracy and inclusiveness. Disagreements were solved with the mediation of TW. The identified barriers and facilitators to harmonized health data collection, sharing and linkage were subsequently clustered into main categories, which were then sub-clustered into smaller categories to highlight the most common barriers and facilitators in these main categories (the full clustering/sub-clustering of barriers and facilitators is shown in Table 1). For the purpose of this systematic review, we defined harmonization techniques as methods which would allow the coherent pooling of different data sources, involving health data collected either prospectively, retrospectively or both. Examples include the use of standard case report forms or data dictionaries, a central review of the collected data, training provided to

researchers/stakeholders and leadership role by one of the partners for coordinating data collection, sharing or linkage activities.

Table 1 Clustering of barriers and facilitators to harmonized health data collection, sharing and linkage

Barriers		Facilitators	
Cluster	**Sub-cluster**	**Cluster**	**Sub-cluster**
Ethical	Privacy	**Ethico-Legal**	Ethical approval by REC/IRB
	Respect for Autonomy		Health Data Anonymization
	Other		Informed Consent
Legal	Data Protection Regulations		Patient data access rights
	Divergence in National Legislations for Data Security and Privacy		Confidentiality measures
	Other		Clarity of legislation for health data collection/sharing/linkage
			Official/legal approval of project
			Study according to International laws and regulations
			Legislation allows project without consent or REC approval
			Legislation requires mandatory reporting
			Other
Technical	Lack of Data Standards (data structure and semantics)	**Technical**	Data harmonization techniques
	Data Quality Issues		Data Linkage techniques
	Limited Technical Capabilities		
	Other		Other
Financial	Lack of Funding	**Financial**	Securing funding
	Other		Public-Private partnership
			Other
Political	Mistrust between stakeholders	**Political**	Data Sharing Agreement
	Data Ownership		Building and maintaining stakeholder trust
	Institutional/constitutional organization issues		Data access control
	Other		Health System Structure
			Other
Motivational	Lack of research incentives	**Motivational**	Monetary Incentive
	Stakeholder restricts access for re-use of data as deemed unfit for secondary use		Easing workload through improvement of data collection
	Stakeholder competing interests		Memorandum of understanding to ensure collaboration until end of study
	Other		Other
Sociocultural	Cultural clash for data collection/sharing/linkage	**Sociocultural**	Participant data access control
	Other		Other

Analysis

A narrative synthesis of included publications was carried out [27]. This step involved the categorization of health data collection, sharing and linkage projects based on their national or cross-national dimension, their source of health data, and barriers and facilitators identified in these publications. This step was important to highlight similarities and differences between projects in Denmark and Switzerland. The statistical software, STATA ® version 15.0, was used for the different analyses.

3.4 Results

A total of 1928 papers were initially retrieved from the search engines and electronic bibliographic databases for the period of January 2008 to December 2017. The search was repeated for the period of January 2018 to March 2019 (upon request of the journal) resulting in a total of 425 additional papers. The result of the two searches were combined for each stage

of the PRISMA resulting in an overall total of 2353 papers retrieved for the period of January 2008 to March 2019 (Fig 1). Duplicates (n = 170) were removed either automatically using ENDNOTE X8 or manually after reviewing abstracts and their titles. The remaining 2183 papers went through title and abstract screening, which resulted in the exclusion of 1789 papers. In-depth full-text screening was performed for 394 papers, and 115 more papers were excluded for not meeting the inclusion criteria (see Fig 1 for reasons). Reference screening of the 279 included papers, resulted in the identification and inclusion of 66 additional papers which met the eligibility criteria for this systematic review (Fig 1).

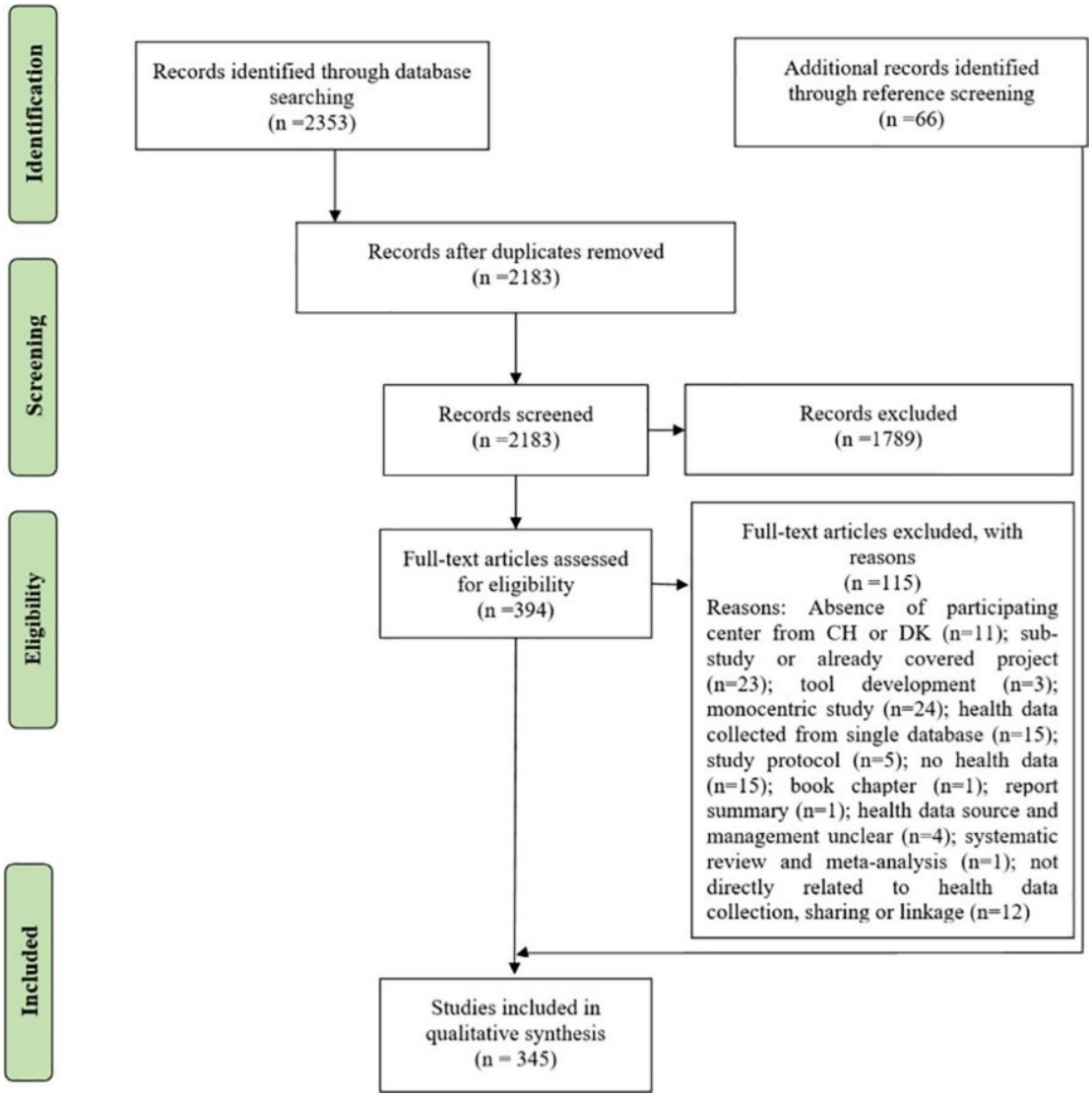


Figure 1 Flow diagram of study selection.

The 345 included papers are summarized in Table 2, where they are categorized based on their national (n = 240) or cross-national (n = 105) dimension, their sources of health data and the total number of barriers and facilitators reported for each project. We identified 200 Danish and 40 Swiss national projects, and 105 cross-national projects. Among these cross-national

projects, 14 projects involved the use of health data from both Denmark and Switzerland, 51 and 40 projects involved a Danish partner and a Swiss partner respectively. Overall, the number of projects which involved primary health data collection, sharing and analysis were 106 (30.7%), 92 (26.7%) and 106 (30.7%) respectively. Comparatively, the number of projects which involved secondary health data collection and analysis were 283 (82.0%; if a study collected both primary health data and secondary health data, it was counted for both). Of the 345 projects, 199 used health data from health services, 211 from public health sector, 94 from research and 62 from other health data sources.

Table 2 Characteristics of included projects (n = 345) with the total number of identified barriers and facilitators per project.

Reference	Country	Partnership Type	Source of health data	Total Identified Barriers, n	Total Identified Facilitators, n
Aabakke et al. 2014 [28]	DK[a]	National	Health Services; Public Health; Other	4	5
Adam et al. 2010 [29]	CH[b]	National	Research	5	7
Agergaard et al. 2017 [30]	DK	National	Health Services	3	6
Agten et al. 2017 [31]	CH	National	Public Health	1	3
Ammundsen et al. 2012 [32]	DK	National	Health Services; Research	3	6
Andersen et al. 2011 [33]	DK	National	Health Services; Public Health	4	8
Andersen et al. 2014 [34]	DK	National	Health Services, Public Health	2	4
Andres et al. 2018 [35]	CH	National	Public Health	4	3
Annaheim et al. 2018 [36]	CH	National	Health Services; Other	5	3
Antonsen et al. 2011 [37]	DK	National	Public Health; Research	2	2
Antonsen et al. 2016 [38]	DK	National	Health Services; Public Health; Research	2	12
Arboe et al. 2016 [39]	DK	National	Health Services; Public Health; Other	0	14
Arking et al. 2014 [40]	CH	Cross-national	Research	0	4
Atlado´ttir et al. 2012 [41]	DK	National	Health Services; Public Health; Other	1	4
Aubert et al. 2016 [42]	CH	National	Health Services	0	5
Auer et al. 2014[43]	CH	National	Health Services; Public Health	0	9
Avillach et al. 2013 [44]	DK	Cross-national	Health Services; Public Health	5	12
Avlund et al. 2018 [45]	DK	National	Health Services; Public Health	1	8
Bachelet et al. 2016 [46]	DK	Cross-national	Health Services; Public Health; Research	2	7
Baker et al. 2009 [47]	DK	National	Health Services	3	2
Baldur-Felskov et al. 2013 [48]	DK	National	Health Services; Research	2	5
Balgobind et al. 2009 [49]	DK	Cross-national	Health Services; Research	0	1
Bay-Nielsen et al. 2008 [50]	DK	National	Public Health	0	2
Beduneau et al. 2017 [51]	CH	Cross-national	Health Services; Public Health	1	6
Begre et al. 2010 [52]	CH	National	Other	2	4
Bendixen et al. 2019 [53]	DK	National	Health Services; Public Health; Research; Other	0	3
Beretta-Piccoli et al. 2017 [54]	CH	National	Health Services; Public Health	3	8
Binderup et al. 2018 [55]	DK	National	Health Services; Research; Other	2	5

Bisgaard et al. 2013 [56]	DK	National	Health Services; Research	0	5
Bjerregaard and Larsen 2011 [57]	DK	National	Health Services; Public Health	0	11
Bjornholt et al. 2015 [58]	DK	National	Health Services; Public Health; Research	1	9
Blaha et al. 2016 [59]	DK	Cross-national	Health Services	2	6
Blenstrup and Knudsen 2011 [60]	DK	National	Health Services; Research	1	3
Blichert-Toft et al. 2008 [61]	DK	National	Health Services; Public Health; Research	0	3
Bodin et al. 2018 [62]	DK	Cross-national	Health Services; Public Health	3	6
Boje et al. 2014 [63]	DK	National	Public Health	1	3
Brenner et al. 2011 [64]	CH	National	Public Health	1	6
Brink et al. 2018 [65]	DK	National	Research	5	6
Burgstaller et al. 2016 [66]	CH	National	Health Services; Research	0	6
Cainzos-Achirica et al. 2018 [67]	DK	Cross-national	Health Services	4	5
Calhaz-Jorge et al. 2017 [68]	DK	Cross-national	Public Health; Other	2	4
Calvet et al. 2014 [69]	CH	Cross-national	Research	2	3
Carstensen et al. 2008 [70]	DK	National	Health Services	0	4
Caspersen et al. 2008 [71]	DK	National	Health Services; Public Health	0	7
Chaigne et al. 2017[72]	CH	National	Health Services	0	5
Chesnaye et al. 2014 [73]	DK	Cross-national	Public Health	1	2
Chmiel et al. 2011 [74]	CH	National	Health Services	3	10
Christensen et al. 2011 [75]	DK	National	Health Services; Public Health; Research	0	2
Christensen et al. 2011b [76]	DK	National	Public Health	0	11
Christensen et al. 2011c [77]	DK	National	Health Services; Public Health	1	5
Christensen et al. 2014 [78]	DK	National	Health Services; Public Health; Other	1	8
Christensen et al. 2016 [79]	DK	National	Public Health	1	6
Christensen et al. 2016b [80]	DK	National	Health Services; Public Health	1	6
Christiansen et al. 2008 [81]	DK	National	Public Health; Research	1	5
Christiansen et al. 2008b [82]	DK	National	Health Services; Public Health	1	3
Christoffersen et al. 2015 [83]	DK	National	Health Services; Public Health; Other	1	5
Coleman et al. 2011 [84]	DK	Cross-national	Public Health	2	13
Coloma et al. 2011 [85]	DK	Cross-national	Health Services	5	9
Corraini et al. 2017 [86]	DK	National	Public Health	1	6
Costantino et al. 2018 [87]	DK	Cross-national	Health Services; Other	4	6
Cotter et al. 2013 [88]	CH	Cross-national	Health Services; Research	1	3
Czauderna et al. 2016 [89]	CH	Cross-national	Research	4	9
Dalgard et al. 2010 [90]	DK	Cross-national	Research	0	0
Damgaard et al. 2013 [91]	DK	National	Public Health	1	3
Darby et al. 2013 [92]	DK	Cross-national	Health Services; Public Health	1	4
Dastani et al. 2012 [93]	CH	Cross-national	Research	0	4
De Angelis et al. 2009 [94]	Both	Cross-national	Research	5	7
De Groot et al. 2014 [95]	DK	Cross-national	Health Services; Public Health; Research	4	1
della Torre et al. 2012 [96]	CH	National	Health Services	0	5
Dencker et al. 2016 [97]	DK	National	Public Health	1	1

De Vos Andersen et al. 2017 [98]	DK	National	Health Services; Public Health; Other	2	9
Diel et al. 2010 [99]	CH	National	Health Services	3	7
Disanto et al. 2016 [100]	CH	National	Public Health	0	11
Donia et al. 2017 [101]	DK	National	Public Health; Research	2	0
Downs et al. 2016 [102]	DK	Cross-national	Public Health; Other	1	2
Dreyer et al. 2015 [103]	DK	Cross-national	Public Health; Other	2	7
Edgren et al. 2015 [104]	DK	Cross-national	Health Services; Public Health; Other	1	6
Ehlers et al. 2009 [105]	DK	National	Public Health; Other	0	3
Ekelund et al. 2015 [106]	DK	National	Health Services; Public Health	1	12
El-Galaly et al. 2015 [107]	DK	Cross-national	Public Health	1	5
Elliott et al. 2017 [108]	DK	National	Health Services	1	2
Engelberger et al. 2015 [109]	CH	National	Health Services; Public Health	0	6
Erdem et al. 2015 [110]	DK	Cross-national	Health Services; Other	1	2
Erichsen et al. 2010 [111]	DK	National	Health Services; Public Health; Research	0	13
Erichsen et al. 2011 [112]	DK	National	Health Services; Public Health	1	4
Erlangsen et al. 2008 [113]	DK	National	Public Health	1	5
Escala-Garcia et al. 2019 [114]	DK	Cross-national	Research	2	2
Escott-Price et al. 2014 [115]	CH	Cross-national	Research	0	0
Fago¨-Olsen et al. 2012 [116]	DK	National	Public Health	0	1
Fahrner et al. 2014 [117]	CH	National	Health Services	4	2
Fedder et al. 2013 [118]	DK	National	Public Health	2	5
Fenger et al. 2016 [119]	DK	National	Public Health; Other	1	5
Fieten et al. 2018 [120]	CH	Cross-national	Research	3	3
Fløe et al. 2018 [121]	DK	National	Health Services; Other	1	4
Frandsen et al. 2014 [122]	DK	National	Health Services; Public Health	2	5
Frary et al. 2016 [123]	DK	National	Health Services; Public Health	0	5
Freiberg et al. 2017 [124]	Both	Cross-national	Health Services; Public Health	0	3
Friis et al. 2009 [125]	DK	National	Health Services; Public Health; Other	1	4
Funcke et al. 2016 [126]	CH	Cross-national	Health Services; Public Health	1	7
Furtwa¨ngler et al. 2018 [127]	CH	Cross-national	Health Services; Research	1	4
Gammelager et al. 2012 [128]	DK	National	Health Services; Public Health	1	7
Garcia-Etienne et al. 2019 [129]	CH	Cross-national	Health Services	1	4
Gatta et al. 2017 [130]	CH	Cross-national	Research	1	1
Gatzioufas et al. 2016 [131]	CH	Cross-national	Research	0	4
Geissbuhler 2013 [132]	CH	National	Health Services	16	15
Ghith et al. 2012 [133]	DK	National	Research; Other	2	7
Gjerstorff 2011 [134]	DK	National	Public Health	1	7
Glintborg et al. 2011 [135]	DK	National	Health Services; Public Health	2	5
Godballe et al. 2009 [136]	DK	National	Public Health	0	5
Gorski et al. 2015 [137]	CH	Cross-national	Research	1	6
Goutaki et al. 2017 [138]	Both	Cross-national	Health Services	2	14

Goutaki et al. 2019 [139]	CH	National	Health Services; Research	4	10
Gradel et al. 2008 [140]	DK	National	Health Services; Public Health	0	5
Grann et al. 2011 [141]	DK	National	Health Services; Public Health	1	6
Gratwohl et al. 2015 [142]	CH	Cross-national	Research	0	4
Gregersen et al. 2016 [143]	DK	National	Public Health	3	6
Griffin et al. 2011 [144]	DK	Cross-national	Health Service; Research; Other	2	3
Gromov et al. 2014 [145]	DK	National	Health Services	0	5
Gruber et al. 2018 [146]	CH	Cross-national	Health Services	0	2
Gudbrandsdottir et al. 2012 [147]	DK	National	Health Services; Other	1	1
Gulmez et al. 2009 [148]	DK	National	Health Services; Public Health	0	4
Gylvin et al. 2017 [149]	DK	National	Health Services; Research; Other	1	3
Hallas et al. 2012 [150]	DK	National	Health Services; Public Health	3	6
Hallas and Pottegard 2017 [151]	DK	National	Health Services; Public Health	1	5
Halmin et al. 2017 [152]	DK	Cross-national	Public Health	0	4
Hansen et al. 2008 [153]	DK	National	Health Services; Public Health	1	8
Hansen et al. 2012 [154]	DK	National	Health Services; Public Health	2	4
Hansen and Jacobsen 2014 [155]	DK	National	Health Services; Research	1	6
Hansen et al. 2018 [156]	DK	National	Health Services; Research; Other	2	4
Harshman et al. 2012 [157]	DK	Cross-national	Health Services; Public Health	1	2
Hatz et al. 2011 [158]	CH	National	Public Health	1	6
Haueis et al. 2012 [159]	CH	Cross-national	Research	1	4
Havelin et al. 2009 [160]	DK	Cross-national	Public Health	3	6
Head et al. 2013 [161]	DK	Cross-national	Health Services; Other	3	5
Helgstrand et al. 2010 [162]	DK	National	Health Services; Public Health	0	7
Helgstrand et al. 2012 [163]	DK	National	Health Services; Public Health; Other	0	3
Helqvist et al. 2012 [164]	DK	National	Health Services; Public Health	0	3
Helweg-Larsen 2011 [165]	DK	National	Public Health; Other	1	3
Hemkens et al. 2017 [166]	CH	National	Research	0	5
Henningsen et al. 2011 [167]	DK	National	Public Health	0	4
Henningsen et al. 2011b [168]	DK	Cross-national	Public Health	4	8
Henriksen et al. 2013 [169]	DK	National	Public Health	0	3
Herzberg et al. 2012 [170]	DK	National	Health Services	1	3
Hetland 2011 [171]	DK	National	Health Services; Other	5	16
Holland-Bill et al. 2014 [172]	DK	National	Health Services; Public Health	1	8
Horsdal et al. 2012 [173]	DK	National	Health Services; Public Health	2	5
Hyldig et al. 2019 [174]	DK	National	Health Services; Public Health; Research; Other	1	5
Ingeholm et al. 2016 [175]	DK	National	Health Services; Public Health; Other	2	6
Ittermann et al. 2018 [176]	DK	Cross-national	Research	1	4
Iversen et al. 2016 [177]	DK	National	Public Health	1	8
Jacobs et al. 2014 [178]	CH	Cross-national	Research	0	11
Jakobsen et al. 2017 [179]	DK	National	Public Health	2	2

Jensen et al. 2009 [180]	DK	National	Health Services; Public Health	1	6
Jensen et al. 2010 [181]	DK	National	Health Services; Public Health	1	0
Jensen et al. 2011 [182]	DK	National	Health Services	2	6
Jensen et al. 2016 [183]	DK	National	Public Health	1	7
Jensen et al. 2017 [184]	DK	National	Public Health	1	1
Jeppesen et al. 2016 [185]	DK	National	Health Services; Public Health	2	5
Johannesdottir et al. 2012 [186]	DK	National	Health Services; Public Health	3	9
Jørgensen et al. 2018 [187]	DK	National	Health Services; Public Health	1	6
Joshi et al. 2015 [188]	Both	Cross-national	Research	1	6
Kachuri et al. 2018 [189]	DK	Cross-national	Research	0	2
Kaltoft et al. 2009 [190]	DK	National	Health Services; Public Health	2	3
Karkov et al. 2010 [191]	DK	National	Health Services; Public Health; Other	2	5
Kent et al. 2015 [192]	DK	National	Health Services; Public Health; Other	1	13
Khanna et al. 2008 [193]	CH	National	Research	1	1
Khatami et al. 2016 [194]	Both	Cross-national	Health Services	2	14
Kiderlen et al. 2012 [195]	CH	Cross-national	Public Health	3	2
Kildemoes et al. 2011 [196]	DK	National	Health Services; Public Health	1	8
Kirwan et al. 2008 [197]	Both	Cross-national	Research	2	12
Klein et al. 2012 [198]	DK	National	Health Services; Public Health	1	1
Knudsen et al. 2013 [199]	DK	National	Health Services	0	1
Kowalska et al. 2011 [200]	DK	Cross-national	Health Services; Other	4	4
Kronborg et al. 2009 [201]	DK	National	Public Health; Other	1	6
Laenkholm et al. 2018 [202]	DK	National	Health Services; Public Health	0	8
Laguna et al. 2009 [203]	CH	Cross-national	Health Services	0	3
Landolt et al. 2016 [204]	CH	Cross-national	Research	0	3
Lang et al. 2019 [205]	CH	Cross-national	Research	2	3
Lange et al. 2017 [206]	DK	National	Health Services; Other	0	5
Laouali et al. 2018 [207]	DK	Cross-national	Public Health; Other	1	6
Larsen et al. 2016 [208]	DK	National	Public Health; Research	2	3
Larsen et al. 2016b [209]	DK	National	Health Services; Public Health	3	5
Laursen et al. 2018 [210]	DK	National	Health Services; Public Health; Research	1	5
Leboeuf-Yde et al. 2012 [211]	DK	National	Research; Other	1	1
Lehnert et al. 2018 [212]	DK	National	Public Health	1	3
Lildballe et al. 2014 [213]	DK	National	Health Services; Public Health	0	2
Linauskas et al. 2018 [214]	DK	National	Public Health	7	4
Lindhardsen et al. 2011 [215]	DK	National	Health Services	1	7
Lindhardsen et al. 2012 [216]	DK	National	Health Services; Other	2	8
Linnet et al. 2009 [217]	DK	National	Health Services; Public Health; Other	2	7
Liu et al. 2016 [218]	DK	National	Health Services; Public Health; Research	1	7
Lund et al. 2018 [219]	DK	National	Public Health	4	8
Lundstrøm et al. 2009 [220]	DK	National	Public Health	3	5

Luta et al. 2018 [221]	CH	National	Research	0	6
Lydiksen et al. 2014 [222]	DK	National	Health Services; Public Health	0	3
Lynge et al. 2011 [223]	DK	National	Health Services	3	4
Maeng et al. 2008 [224]	DK	National	Health Services; Public Health	1	4
Mahajan et al. 2018 [225]	DK	Cross-national	Research	2	6
Majholm et al. 2012 [226]	DK	National	Health Services; Public Health	3	3
Mareri et al. 2011 [227]	Both	Cross-national	Research	0	4
Margulis et al. 2017 [228]	DK	Cross-national	Public Health	2	5
May et al. 2014 [229]	CH	Cross-national	Research	3	6
Mejdahl et al. 2013 [230]	DK	National	Public Health; Other	2	3
Mellernkjær et al. 2014 [231]	DK	National	Health Services; Public Health	0	1
Messerli et al. 2016 [232]	CH	National	Public Health; Research	0	6
Mikkelsen et al. 2015 [233]	DK	National	Health Services; Other	2	4
Minnerup et al. 2015 [234]	CH	Cross-national	Health Services	0	1
Modvig et al. 2017 [235]	DK	National	Public Health	0	5
Mo¨hring et al. 2019 [236]	CH	Cross-national	Research	1	7
Møller et al. 2008 [237]	DK	National	Public Health	2	6
Mors et al. 2011 [238]	DK	National	Health Services; Public Health	0	9
Mortensen et al. 2011 [239]	DK	National	Public Health; Research	1	3
Mortensen et al. 2013 [240]	DK	National	Health Services; Public Health	0	2
Mueller et al. 2015 [241]	CH	Cross-national	Research	0	7
Mukai et al. 2013 [242]	DK	National	Health Services; Public Health	1	2
Mu¨ller et al. 2012 [243]	CH	National	Other	1	4
Munk et al. 2012 [244]	DK	National	Public Health; Other	0	6
Narath et al. 2016 [245]	CH	Cross-national	Health Services; Research	0	7
Neelon et al. 2015[246]	DK	National	Public Health	1	2
Nickenig et al. 2014 [247]	Both	Cross-national	Health Services; Public Health	2	4
Nielsen et al. 2012 [248]	DK	National	Health Services; Public Health	1	3
Nielsen et al. 2015 [249]	DK	National	Health Services; Public Health	2	2
Nielsen et al. 2015b [250]	DK	National	Health Services; Public Health	2	3
Nielsen and Nordestgaard 2016 [251]	DK	National	Health Services; Public Health; Other	1	3
Nilsson et al. 2014 [252]	DK	National	Health Services; Public Health	3	5
Nolan-Kenney et al. 2019 [253]	CH	Cross-national	Research	4	8
Nørskov et al. 2015 [254]	DK	National	Public Health	1	4
Nørskov et al. 2017 [255]	DK	National	Public Health; Research	1	4
Nyholm et al. 2015 [256]	DK	National	Health Services; Public Health	2	3
Olsen et al. 2008 [257]	DK	National	Health Services; Public Health; Other	3	4
Olsen et al. 2013 [258]	DK	National	Public Health	1	5
Orsted et al. 2011[259]	DK	National	Public Health	2	5
O¨ zcan et al. 2016 [260]	DK	National	Health Services	2	10
Pacurariu et al. 2015 [261]	DK	Cross-national	Public Health	5	2

Pagh et al. 2013 [262]	DK	National	Health Services; Other	1	2
Palnum et al. 2012 [263]	DK	National	Health Services; Public Health; Other	2	6
Pasternak et al. 2014 [264]	DK	National	Health Services; Public Health	1	5
Patadia et al. 2018 [265]	DK	Cross-national	Health Services; Research	1	2
Pattaro et al. 2016 [266]	Both	Cross-national	Research	1	7
Paulsen et al. 2013 [267]	DK	National	Public Health	1	6
Pechmann et al. 2019 [268]	CH	Cross-national	Research	1	12
Pedersen et al. 2010 [269]	DK	Cross-national	Health Services	0	2
Pedersen 2011 [270]	DK	National	Health Services; Public Health	2	6
Pedersen et al. 2011 [271]	DK	National	Health Services; Public Health	0	4
Perera et al. 2018 [272]	DK	Cross-national	Public Health	3	1
Perregaard et al. 2015 [273]	DK	National	Public Health	2	3
Petersen et al. 2018 [274]	DK	National	Public Health	3	2
Petersen et al. 2018b [275]	DK	National	Health Services; Public Health; Other	1	3
Piazza et al. 2010 [276]	CH	Cross-national	Health Services	0	1
Piltoft et al. 2017 [277]	DK	National	Public Health; Other	0	4
Pinborg et al. 2015 [278]	DK	National	Public Health	0	4
Pironi et al. 2017 [279]	DK	Cross-national	Health Services	2	5
Plu¨ss-Suard et al. 2013 [280]	CH	National	Health Services	1	3
Pommergaard et al. 2014 [281]	DK	National	Health Services; Public Health	3	3
Pottegard et al. 2014 [282]	DK	National	Public Health	0	9
Pottegard et al. 2015 [283]	DK	National	Public Health	2	6
Poulsen et al. 2012 [284]	DK	National	Health Services	2	4
Poulsen et al. 2016 [285]	DK	National	Health Services; Public Health	1	4
Poulsen et al. 2018 [286]	DK	National	Health Services; Public Health	1	6
Preston et al. 2014 [287]	DK	National	Public Health	0	5
Prins et al. 2018 [288]	DK	Cross-national	Research	0	6
Pukkala et al. 2009 [289]	DK	Cross-national	Public Health	2	6
Radovanovic and Erne 2010 [290]	CH	National	Health Services	3	12
Ramlau-Hansen et al. 2009 [291]	DK	National	Health Services; Public Health	2	3
Rasmussen et al. 2012 [292]	DK	National	Public Health	1	2
Rasmussen and Tønnesen 2016 [293]	DK	National	Public Health; Other	1	7
Rasmussen et al. 2017 [294]	DK	National	Public Health; Other	2	7
Rathe 2015 [295]	DK	National	Health Services; Public Health	0	7
Reyes et al. 2016 [296]	DK	Cross-national	Public Health	0	4
Ringdal et al. 2011 [297]	Both	Cross-national	Research; Other	6	8
Roberto et al. 2016 [298]	DK	Cross-national	Health Services; Public Health; Research	2	8
Rudin et al. 2008 [299]	CH	National	Research	0	5
Rungby et al. 2017 [300]	DK	National	Health Services; Public Health	2	6
Russell et al. 2018 [301]	DK	Cross-national	Research	2	5
Schaefer et al. 2013 [302]	CH	National	Health Services	0	4

Scha¨fer et al. 2018 [303]	CH	Cross-national	Research; Other	2	6
Schatlo et al. 2012 [304]	CH	National	Health Services; Research	0	4
Schatorje´ et al. 2014 [305]	CH	Cross-national	Research	4	6
Schmaal et al. 2017 [306]	CH	Cross-national	Health Services	1	6
Schmidt et al. 2010 [307]	DK	National	Public Health; Other	0	4
Schmidt et al. 2010b [308]	DK	National	Public Health	0	8
Schmidt et al. 2011 [309]	DK	National	Public Health	0	4
Schmidt et al. 2012 [310]	DK	National	Public Health	0	5
Schmidt et al. 2012b [311]	DK	National	Public Health	2	5
Schmidt et al. 2014 [312]	DK	National	Public Health; Other	1	11
Schmidt et al. 2018 [313]	DK	National	Health Services	1	11
Schneeberger et al. 2013 [314]	Both	Cross-national	Health Services	0	3
Schoos et al. 2015 [315]	DK	National	Public Health; Other	0	4
Schroll et al. 2012 [316]	DK	National	Health Services	1	4
Schuemie et al. 2012 [317]	DK	Cross-national	Health Services; Public Health	2	6
Sejbaek et al. 2013 [318]	DK	National	Public Health; Research	0	4
Skyum et al. 2018 [319]	DK	National	Health Services; Other	3	4
Skyum et al. 2019 [320]	DK	Cross-national	Health Services; Research	2	7
Soerensen et al. 2014 [321]	DK	National	Health Services; Public Health	1	7
Sommer et al. 2018 [322]	CH	National	Research	2	6
Sørensen et al. 2009 [323]	DK	National	Public Health	1	1
Sørensen et al. 2013 [324]	DK	National	Health Services; Public Health	1	6
Spoerri et al. 2010 [325]	CH	National	Public Health	1	3
Stahl Madsen et al. 2014 [326]	DK	National	Health Services	0	3
Steenholdt et al. 2014 [327]	DK	National	Public Health; Research	1	6
Stewardson et al. 2016 [328]	CH	Cross-national	Health Services	1	9
Strasser et al. 2016 [329]	CH	National	Research	1	7
Streit et al. 2014 [330]	CH	National	Health Services	2	4
Strnad et al. 2016 [331]	CH	Cross-national	Research	0	6
Stukalin et al. 2018 [332]	DK	Cross-national	Health Services; Public Health	1	3
Su¨rder et al. 2013 [333]	CH	National	Research	1	4
Suttorp et al. 2018 [334]	CH	Cross-national	Research	0	6
Svendsen et al. 2013 [335]	DK	National	Health Services; Public Health	0	2
Talman et al. 2008 [336]	DK	National	Health Services; Public Health	0	3
Thillemann et al. 2009 [337]	DK	National	Public Health	0	4
Thomsen et al. 2008 [338]	DK	National	Health Services; Public Health	0	5
Thornqvist et al. 2014 [339]	DK	National	Health Services; Public Health	2	6
Thøstesen et al. 2015 [340]	DK	National	Public Health; Research; Other	0	4
Thygesen et al. 2011 [341]	DK	National	Health Services; Public Health	0	5
Tollånes et al. 2016 [342]	DK	Cross-national	Health Services; Public Health; Research	2	7
Trabert et al. 2014 [343]	DK	Cross-national	Research	4	4

Tutolo et al. 2019 [344]	CH	Cross-national	Health Services	1	3
Tvedskov et al. 2011 [345]	DK	National	Health Services; Public Health	2	5
Tvedskov et al. 2015 [346]	DK	National	Health Services; Public Health	1	5
Ulff-Moller et al. 2018 [347]	DK	National	Health Services; Public Health; Research; Other	2	6
Underbjerg et al. 2013 [348]	DK	National	Health Services; Public Health	2	6
Underbjerg et al. 2015 [349]	DK	National	Public Health	0	5
Ungaro et al. 2019 [350]	DK	Cross-national	Public Health	3	6
Usvyat et al. 2013 [351]	Both	Cross-national	Health Services	8	4
Vach et al. 2018 [352]	CH	National	Health Services; Research; Other	2	4
Van Hedel et al. 2018 [353]	CH	Cross-national	Health Services	3	7
Van Stralen et al. 2011 [354]	Both	Cross-national	Research	1	3
Vasan et al. 2016 [355]	DK	Cross-national	Public Health	0	5
Vester-Andersen et al. 2014 [356]	DK	National	Health Services; Public Health	0	6
Vest-Hansen et al. 2014 [357]	DK	National	Public Health	1	5
Viberg et al. 2018 [358]	DK	National	Health Services; Public Health	0	5
Villadsen et al. 2011 [359]	DK	National	Health Services; Public Health	2	3
Walters et al. 2013 [360]	DK	Cross-national	Public Health	5	4
Weber et al. 2013 [361]	CH	National	Health Services	11	5
Weigang et al. 2010 [362]	CH	Cross-national	Health Services	1	3
Wiegand et al. 2014 [363]	CH	Cross-national	Research	1	6
Wildgaard et al. 2011 [364]	DK	National	Health Services; Public Health	1	2
Winterfeld et al. 2013 [365]	Both	Cross-national	Health Services	1	4
Wurtzen et al. 2013 [366]	DK	National	Health Services; Research	0	6
Ylijoki-Sorensen et al. 2014 [367]	DK	Cross-national	Public Health	4	4
Zalfani et al. 2012 [368]	CH	National	Health Services; Public Health	0	4
Zecca et al. 2018 [369]	CH	National	Health Services	2	4
Zellweger et al. 2014 [370]	CH	National	Health Services; Other	2	5
Zellweger et al. 2019 [371]	CH	National	Public Health	1	5
Zwisler et al. 2016 [372]	DK	National	Health Services; Public Health	0	10

[a] DK: Denmark [b] CH: Switzerland

Overview of barriers

Barriers of an ethical nature were reported 19 times in the included records and they concerned mainly issues related to privacy (n = 9) and respect for autonomy of study participants (n = 6) (Table 3). As to legal barriers, these were reported 17 times and they included issues associated with national data protection regulations (n = 4), differences in national legislations concerning data security and privacy (n = 4) and "Other" (n = 9) (e.g. legal uncertainty concerning health data collection or sharing, market restriction, etc.). Overall, the type of barriers that were more often reported, however, were those of a technical nature. In the records, 416 technical barriers were mentioned and they were classified as data quality issues (e.g. data incompleteness,

potential misclassification of data, etc.) (n = 234), lack of data standards (data structure and semantics, e.g. ambiguous terminologies, temporal evolution of data standards, etc.) (n = 151), limited technical capabilities (e.g. no unique identifier, etc.) (n = 21) and "Other" (n = 10) (e.g. time constraints on physicians preventing the use of standard procedures for data collection). Financial barriers were also reported, but only a limited amount of times (n = 9), and they were principally referring to the unavailability or inadequacy of financial support (n = 8). Only 13 political barriers were found and they comprised institutional/constitutional organization issues (e.g. federalist system and different healthcare systems) (n = 6), mistrust between stakeholders (n = 3), data ownership issues (n = 2) and "Other" (n = 2) (e.g. no official guidelines for data sharing). Studies also reported some motivational barriers, including lack of research incentives (n = 17) (including additional workload imposed on physicians/researchers), data re-use prevented by stakeholders as they are deemed unfit for secondary use (n = 2), stakeholders' competing interests (n = 2) and additional barriers of a diversified content, thus labelled as "Other" (n = 4) (e.g. study participants not showing up for part of the study). Finally, 6 socio-cultural barriers were reported in the included records, half of which were related to a "cultural clash" (n = 3), which we defined as issues resulting from different cultures in data collection, sharing and linkage of the partners involved in the project.

Table 3 Distribution of barriers' sub-clusters in national and cross-national Danish and Swiss projects.

Barriers		Countries involved in projects		
Cluster	**Sub-cluster**	**Denmark N[a] = 251**	**Switzerland N = 80**	**Both countries N = 14**
		n[b] (mean no. of barriers per project)	**n (mean no. of barriers per project)**	**n (mean no. of barriers per project)**
Ethical	Privacy	6 (0.02)	3 (0.04)	-[c] (N/A)
	Respect for Autonomy	3 (0.01)	3 (0.04)	- (N/A)
	Other	3 (0.01)	1 (0.01)	- (N/A)
Legal	Data Protection Regulations	2 (0.01)	1 (0.01)	1 (0.07)
	Divergence in National Legislations for Data Security and Privacy	2 (0.01)	- (N/A)	2 (0.14)
	Other	5 (0.02)	3 (0.04)	1 (0.07)
Technical	Lack of Data Standards	104 (0.41)	33 (0.41)	14 (1.00)
	Data Quality Issues	181 (0.72)	44 (0.55)	9 (0.64)
	Limited Technical Capabilities	11 (0.04)	9 (0.11)	1(0.07)
	Other	8 (0.03)	2 (0.03)	- (N/A)
Financial	Lack of Funding	4 (0.02)	3 (0.04)	1 (0.07)
	Other	1 (0.00)	- (N/A)	- (N/A)
Political	Mistrust between stakeholders	- (N/A)	3 (0.04)	- (N/A)
	Data Ownership	2 (0.01)	- (N/A)	- (N/A)
	Institutional/constitutional organization issues	2 (0.01)	4 (0.05)	- (N/A)
	Other	- (N/A)	2 (0.03)	- (N/A)
Motivational	Lack of research incentives	6 (0.02)	9 (0.11)	2 (0.14)
	Stakeholder restricts access for re-use of data as deemed unfit for secondary use	2 (0.01)	- (N/A)	- (N/A)
	Stakeholder competing interests	1 (0.00)	1 (0.01)	- (N/A)
	Other	1 (0.00)	3 (0.04)	- (N/A)
Sociocultural	Cultural clash for data collection/sharing/linkage	1 (0.00)	2 (0.03)	- (N/A)
	Other	1 (0.00)	2 (0.03)	- (N/A)

[a] N is the total number of projects in each country category
[b] n is the total number of reported barriers per sub-cluster
[c] –is the absence of reported barriers per sub-cluster
N/A–Not Applicable

Overview of facilitators

Facilitators of an ethico-legal nature were reported 582 times in total, and they were classified as official/legal approval of study (e.g. Danish Data Protection Agency) (n = 148), ethical approval by a REC/IRB (n = 135), legislation permitting to proceed with health data collection, sharing and linkage without consent or REC/IRB approval (n = 79), obtaining informed consent from participants (n = 69), health data anonymization (n = 58), the presence of legislation requiring mandatory reporting (n = 41), confidentiality measures (n = 29; e.g. data security audits), project done according to international laws and regulations (n = 8), data access rights for patients (n = 4), clear legislation for data collection, sharing or linkage (n = 3) and "Other" (n = 8) (e.g. study data made available by researchers upon request). Facilitators of a technical nature were reported 981 times in total, which were grouped in three categories, namely techniques for data harmonization (n = 798), data linkage (n = 155) and "Other" (n = 28) (e.g. study allowed the creation of optional and mandatory datasets, whereby a minimum of data are classified as mandatory). Facilitators of a financial nature, especially explaining how funding was successfully secured, were mentioned 12 times. These referred, for example, to public-private partnerships, where both partners would gain some benefits from the collaboration, as a solution for funding issues (n = 3). 169 facilitators related to politics were reported. These referred to the structure of the health system as an advantage for harmonized health data collection, sharing and linkage (n = 139), data access control by the players (n = 11), the presence of a data sharing agreement between the stakeholders (n = 9), building and maintaining stakeholders' trust for collaboration (n = 7) and "other" (n = 4). There were 14 motivational facilitators, which included monetary incentives to incite researchers/stakeholders to abide by standardized procedures for data handling and management (n = 7), improved data collection tool to ease the workload of researchers/stakeholders for data collection/sharing (n = 3), a memorandum of understanding between partners to ensure collaboration till end of study (n = 2) and "other" (n = 2). Lastly, there were 8 socio-cultural facilitators, which included data subjects controlling access to their data (n = 4) and "Other" (n = 4) (e.g. transparent policies for the participants). Country-wise distribution for all six facilitators categories are presented in Table 4.

Table 4 Distribution of facilitators' sub-clusters in national and cross-national Danish and Swiss projects.

Facilitators		Countries involved in projects		
Cluster	Sub-cluster	Denmark N[a] = 251	Switzerland N = 80	Both countries N = 14
		n[b] (mean no. of facilitators per project)	n (mean no. of facilitators per project)	n (mean no. of facilitators per project)
Ethico-Legal[c]	Ethical approval by REC/IRB	73 (0.29)	55 (0.69)	7 (0.50)
	Health Data Anonymization	31 (0.12)	22 (0.28)	5 (0.36)
	Obtaining informed Consent	29 (0.12)	34 (0.43)	6 (0.43)
	Patient data access rights	3 (0.01)	1 (0.01)	-[d] (N/A)
	Confidentiality measures taken	22 (0.09)	6 (0.08)	1 (0.07)
	Clarity of legislation for health data collection/ sharing/linkage	2 (0.01)	1 (0.01)	- (N/A)
	Official/legal approval of project	140 (0.56)	7 (0.09)	1 (0.07)
	Project done according to international laws and regulations	6 (0.02)	1 (0.01)	1 (0.07)
	Legislation allows project without consent or REC approval	66 (0.26)	12 (0.15)	1 (0.07)
	Legislation requires mandatory reporting	40 (0.16)	1 (0.01)	- (N/A)
	Other	6 (0.02)	2 (0.03)	- (N/A)
Technical	Data harmonization techniques	488 (1.94)	251 (3.14)	59 (4.21)
	Data Linkage techniques	146 (0.58)	6 (0.08)	3 (0.21)
	Other	24 (0.10)	3 (0.04)	1 (0.07)
Financial	Securing funding	6 (0.02)	1 (0.01)	1 (0.07)
	Public-Private partnership	1 (0.00)	2 (0.03)	- (N/A)
	Other	1 (0.00)	- (N/A)	- (N/A)
Political	Data Sharing Agreement	1 (0.00)	5 (0.06)	3 (0.21)
	Building and maintaining stakeholder trust	1 (0.00)	4 (0.05)	2 (0.14)
	Data access control	9 (0.04)	2 (0.03)	- (N/A)
	Health System Structure	138 (0.55)	1 (0.01)	- (N/A)
	Other	3 (0.01)	- (N/A)	- (N/A)
Motivational	Monetary Incentive	5 (0.02)	2 (0.03)	- (N/A)
	Easing workload through improvement of data collection	1 (0.00)	2 (0.03)	- (N/A)
	Memorandum of understanding to ensure collaboration until end of study	- (N/A)	2 (0.03)	- (N/A)
	Other	- (N/A)	1 (0.01)	1 (0.07)
Sociocultural	Participant data access control	2 (0.01)	1 (0.01)	1 (0.07)
	Other	4 (0.02)	- (N/A)	- (N/A)

Table 4 shows the distribution of facilitators' sub-clusters in national and cross-national Danish and Swiss projects. As such, single-country and multi-national countries are not differentiated.
[a]N is the total number of projects in each country category
[b] n is the total number of reported facilitators per sub-cluster
[c]Ethical and legal facilitators were merged as reported solutions had both an ethical and a legal dimension
[d]–is the absence of reported facilitators per sub-cluster
N/A–Not Applicable

Barriers and facilitators identified in national Danish and Swiss projects

When considering only national projects (n = 240) involving either Denmark (N = 200) or Switzerland (N = 40) alone, there were 323 identified barriers and 1234 facilitators. Technical barriers and facilitators were most frequently reported. For comparison purposes and compensation for the imbalances in the number of national projects identified in each country, the absolute numbers and the number of barriers and facilitators per 1,000 national projects for each country are illustrated in Table 5.

Table 5 Distribution of barriers and facilitators in national Danish and Swiss projects.

Barrier category	Denmark N[a] = 200	Switzerland N = 40	Facilitator category	Denmark N = 200	Switzerland N = 40
	n[b] (no. of barriers per 1,000 projects)	n (no. of barriers per 1,000 projects)		n (no. of facilitators per 1,000 projects)	n (no. of facilitators per 1,000 projects)
Ethical	6 (30)	6 (150)	Ethico-legal	331 (1655)	82 (2050)
Legal	6 (30)	4 (100)			
Technical	216 (1080)	51 (1275)	Technical	523 (2615)	132 (3300)
Financial	3 (15)	2 (50)	Financial	8 (40)	2 (40)
Political	-[c] (N/A)	8 (200)	Political	134 (670)	6 (150)
Motivational	7 (35)	8 (200)	Motivational	6 (30)	5 (125)
Sociocultural	2 (10)	4 (100)	Sociocultural	4 (20)	1 (25)
Total	240	83	Total	1006	228
Mean	1.20	2.08	Mean	5.03	5.70

[a] N is the total number of projects in each country category

[b] n is the total number of identified barriers or facilitators per cluster

[c] –is the absence of identified barriers and facilitators per cluster

N/A–Not Applicable

Interestingly, the only identified category of barriers which was comparatively almost equally reported in Swiss and Danish single-country projects was that of technical barriers. Otherwise, ethical, legal, financial, motivational and socio-cultural barriers were reported 5.0, 3.3, 3.3, 5.7 and 10.0 times more in Swiss projects than in Danish projects respectively. On the contrary, a Swiss project reported on average more facilitators than a Danish one (only financial facilitators were reported equally in both countries). Ethico-legal, technical, motivational and socio-cultural facilitators were reported 1.2, 1.3, 4.2 and 1.3 times more in Swiss projects than in Danish projects respectively. Only facilitators related to politics were reported 4.5 times more in Danish projects than Swiss projects.

Barriers and facilitators identified in cross-national Danish and Swiss projects

With respect to cross-national projects (n = 105), there were 182 identified barriers and 532 identified facilitators. Technical barriers and facilitators were more frequently reported than those of another nature. For comparison purposes and compensation for the imbalances in the number of cross-national projects involving each country, the number of barriers and facilitators per 1,000 cross-national projects was calculated (excluding cross-national projects involving both countries) and illustrated in Table 6.

Table 6 Distribution of barriers and facilitators in cross-national Danish and Swiss projects.

Barrier category	Denmark N[a] = 51	Switzerland N = 40	Both countries N = 14	Facilitator category	Denmark N = 51	Switzerland N = 40	Both countries N = 14
	n[b] (Number of barriers per 1,000 projects)	n (Number of barriers per 1,000 projects)			n (Number of facilitators per 1,000 projects)	n (Number of facilitators per 1,000 projects)	
Ethical	6 (118)	1 (25)	-[c]	Ethico-legal	87 (1706)	60 (1500)	22
Legal	3 (59)	- (N/A)	4				
Technical	88 (1725)	37 (925)	24	Technical	135 (2647)	128 (3200)	63
Financial	2 (39)	1 (25)	1	Financial	- (N/A)	1 (25)	1
Political	4 (78)	1 (25)	-	Political	18 (353)	6 (150)	5
Motivational	3 (59)	5 (125)	2	Motivational	- (N/A)	2 (50)	1
Sociocultural	- (N/A)	- (N/A)	-	Sociocultural	2 (39)	- (N/A)	1
Total	106	45	31	Total	242	197	93
Mean	2.08	1.13	2.21	Mean	4.75	4.93	6.64

[a] N is the total number of projects in each country category
[b] n is the total number of identified barriers or facilitators per cluster
c–is the absence of identified barriers and facilitators per cluster
N/A–Not Applicable

Concerning cross-national projects, we observed a reverse tendency as compared to national projects. Studies involving a collaboration with a Swiss partner have, on average, reported 1.8 times less barriers than those involving a Danish partner. More in detail, projects including Switzerland reported 4.7, 1.9, 1.6, 3.1 times less barriers of an ethical, technical, financial and political nature respectively, than those with a Danish partner. However, cross-national projects involving a Swiss partner, reported 2.1 times more barriers of a motivational nature than those with a Danish partner. Comparatively, cross-national collaboration involving either a Swiss or Danish partner reported almost the same number of facilitators. Ethico-legal and political facilitators were identified 1.1 and 2.4 times more in cross-national projects with a Danish partner as opposed to cross-national projects involving a Swiss one. However, technical facilitators were identified 1.2 times more in cross-national projects with a Swiss partner than in those with a Danish one.

3.5 Discussion

This systematic review provides a comprehensive overview of projects from either Denmark or Switzerland which involved the collection, linking or sharing of data and of the barriers and facilitators related to the usage of health data therein reported. Our study includes a broad range of projects relying on data from different sources and contexts (health services, public health, research and other) and it confirms that studies involving the harmonization, linking or sharing of health data still encounter a high number of obstacles, but also underscores that barriers have prompted the development of numerous solutions. We will here address and discuss the findings related to barriers and facilitators of each cluster that was identified.

Ethico-legal barriers and facilitators

Although ethico-legal factors are often described as some of the most problematic elements when it comes to linking and sharing health-related data [9, 373, 374], our results show that barriers of this nature are rarely reported. The small amount of ethico-legal barriers identified might either mean that such barriers were rarely present or that they were present but underreported. In our view, the latter option is more probable for at least two reasons. Firstly, as the records included in this review were all published articles, the explicit mentioning of ethico-legal complications might have been avoided to bypass problems related to publication. Secondly, ethico-legal factors are often less tangibile and transparent in comparison–for example–with technical ones [15] and they are thus more likely to be superseded. Moreover, underreporting would confirm that ethico-legal aspects related to processing of health data are still underappreciated, which is a major obstacle to the final success of research projects [2]. This also suggests that there is some resistency by authors to openly disclose and discuss ethico-legal problematics. For the future, a less cautious approach would be much more beneficial, since it would allow new research projects to build on the issues encountered by old ones.

Ethico-legal facilitators were more widely mentioned. Results show that Swiss projects are still predominantly anchored to the "consent or anonymise" approach, according to which the solution to solve ethico-legal problematics concerning health data is to either anonymize information or to require explicit authorization by data subjects [1]. Differently, Danish projects have made vaster use of alternative solutions, such as relying on specific confidentiality tools, and, more importantly, exploiting regulation that allows—upon certain conditions—to share and link health-related data without the need of obtaining consent by data subjects or REC approval. This demonstrates that the development of proper regulations to facilitate the harmonization and linking of health data offers practical solutions that projects developers are then willing to use. In this framework, another important finding concerns the role of the data protection authority. Whereas in Switzerland this public office–although existing–does not play a defined role with respect to research, results show that Danish studies have a more active interaction with the Data Protection Agency, as they need to apply for permission to use health data. The nature of the application to the national Data Protection Agency that Danish projects need to file is not explicitly described in the records reviewed, but it has been presented elsewhere [375, 376] as a less demanding procedure, resembling a simple duty of notification. Thus, many Danish projects dealing exclusively with health data–in accordance with national regulation–do not need to apply for full ethical review from a REC or IRB, an often demanding and lengthy process, but simply have to obtain clearance from the Data Protection Agency. This

institutionalized interaction with the public authority responsible to ensure compliance with data processing rules can be an important factor helping project developers, since it incentivizes to proactively tackle privacy concerns. This interaction could thus be considered as a model to inspire changes in the regulatory framework in Switzerland.

Technical barriers and facilitators

In this systematic review, data quality issues were the most commonly reported barriers, followed by the lack of data standards and limited technical capabilities. Although Denmark has a developed health data infrastructure, numerous identified projects described that data quality problems still affect health services, public health and research datasets [38, 79, 86, 98, 119, 143, 149, 151]. This is confirmed by other studies, such as a review on the Danish National Patient Registry (DNRP) where the authors concluded that data incompleteness and heterogeneous validation methods of data limited the research potential of this registry [377]. Although relevant, data quality issues can be mitigated in a system like the Danish one, since linkage between data from different registries can be easily performed using the personal identification number (CPR) provided to all Danish citizens at birth and to stable residents [270]. Comparatively, Swiss projects and projects involving a Swiss partner also reported slightly more issues related to data quality than to data standards. However, in comparison to their Danish counterparts which reported almost twice more issues related to data quality than data standards, the difference in reporting of data standards and data quality issues was smaller in Swiss projects. This more equivalent reporting could imply that data standard issues are considered as important as data quality issues for the success of Swiss projects. Indeed, the high levels of data-heterogeneity in the Swiss healthcare context might stem from the fragmented nature of the healthcare system, where each of the 26 cantons [federal states] has a high degree of autonomy and where more than 55 health insurers are active [378].

These findings underline how technical issues are interconnected with the context where projects are carried out, and that also external systemic factors–and not simply internal complications of the projects themselves—affect the emerging of these barriers. In Denmark, for example, the presence of nation-wide registries fosters the development of studies relying on secondary use of routinely collected data, where researchers are more likely faced with issues about the quality of data, since the latter was originally collected for a different purpose. On the contrary, in a country like Switzerland—where data are more often prospectively collected—issues about the absence of common standards because of fragmentation are also likely to be evident, on top of data quality issues.

Our findings suggest therefore that even technical issues concerning data are strongly embedded in the surrounding where projects are conceived. This should induce project developers to communicate and learn from each others, since the barriers they will encounter and the solutions they will find are more likely to be dependent also on the context where they act, and not only on the specific features of their research. For example, since Switzerland's healthcare sector does not use a universal personal identification number because of privacy concerns [379], linkage of data will almost certainly represent a technical challenge, regardless of the features the single project or the data that it aims at using.

Motivational and financial barriers and facilitators

With respect to motivational and financial factors, our findings are partly in line with the literature. Previous research had underscored that the key motivational and financial aspects concerned the lack of research incentives from resource-limited institutions, the fears of being 'robbed' of data before publication or of losing reputation because others might identify errors in the data, the reluctance to facilitate access due to potential inappropriateness of further uses, the need to secure resources for data sharing activities and the necessity to make arrangements between institutions for data management costs [15, 380, 381].

Overall, national and cross-national Swiss projects combined reported more frequently motivational and financial facilitators than their Danish counterparts. This suggests that in a country with a less institutionalised system of data sharing and where studies often have a prospective design, more strategies are elaborated to deal with financial and motivational issues related to data, since–with a lower systemic support–single project developers have to make a greater effort. In contrary, in a context like Denmark—with the high prevalence of studies with retrospective design and the reliance on secondary uses of routinely collected health data–the need for financial and motivational facilitators might be lower. In fact, when health data harmonization is prevalently retrospective, a lower number of actors is involved [382]–since primary data collectors are rarely included–thus reducing the urgency to create motivational or financial incentives for a large number of collaborators.

Another important finding related to financial aspects is that the presence of economic constraints can be the source of additional barriers related to data harmonization, such as data quality issues. For instance, the Swiss project AMIS Plus—concerning a register for acute coronary syndrome—could not envisage systematic site visits to assess data quality or more in-depth questionnaires due to resource limitation [290]. In Denmark, similarly, with the Copenhagen School Health Records Register—a health examination register for schoolchildren

containing data on more than 350,000 individuals—financial constraints made it impossible for the authors to computerize the entire health card, thus limiting the understanding of potential confounding variables [47]. This indicates even more that barriers of different natures are interconnected and that new projects need to acknowledge this interconnectedness of the barriers to successfully address them.

Political barriers and facilitators

Danish national projects did not report any barriers of a political nature, whereas cross-national collaborations mentioned a few, such as data ownership and organizational issues [44, 85, 95]. This suggests that an institutionalization of data processing practices, similar to what occurs in Denmark [383], helps to remove political obstacles. Moreover, the presence of a centralized healthcare system structure also proves helpful, because it reduces the number of actors involved and thus the presence of competing interests. Political issues, however, might re-emerge when projects are cross-national and thus abandon the relatively *safe-haven* created at the national level.

In a context like the Swiss one, on the contrary, political barriers seem to be more relevant for national projects, because these fuel internal conflicts related to the diversity of interests within healthcare and to the difficulty of implementing uniform and centralized policies [132]. In fact, the two most mentioned political facilitators in Switzerland–building trust amongst stakeholders [132, 361] and stakeholders retaining control over data access [132, 290]–are both related to the attempt to coordinate the numerous different parties operating in the health data field and accommodate their competing interests. This might also explain why less political barriers are reported for Swiss cross-national projects. In fact, when projects from a context like the Swiss one go to a supra-national level, the chances of disputes related to in-country political antagonism to emerge is lower.

Our results are thus in line with the literature, where mistrust between stakeholders, absence of comprehensive guidelines for data sharing and lack of legal accountability were identified as major political issues [2, 7, 15]. However, our results further show that the incidence of political barriers seems quite different in single-country studies if compared to cross national ones. This finding is particularly important since it underlines that sometimes the choice of a national or cross-national design might have an impact on the number of political issues encountered.

Socio-cultural barriers and facilitators

Barriers and facilitators of a socio-cultural nature were rarely mentioned in the included records. Comparatively, the incidence of cultural barriers seems to be higher for Switzerland, where

cultural clashes were mentioned more often than for Danish projects. Such difference could be due to the higher degree of fragmentation of the Swiss healthcare system in comparison to the Danish one, which is centralized and state-funded [18]. In fact, one Swiss study [132] reported that the choice for a distributed model in the managing of data was based on prior failures to implement centralized systems of health data and public mistrust towards the concept of centralization. Socio-cultural facilitators were mostly related to the involvement of data-subjects by allowing them to retain control of data access. For instance, the Swiss project reported that data subjects had the possibility to decide which part of their medical records could be considered "stigmatizing", and thereafter blinded to healthcare professionals, other than their designated and trusted physician [132]. The designated and trusted physician would have access to the full record.

It is naturally impossible to determine whether socio-cultural barriers were actually overlooked or simply not reported. In either case, the limited mentioning of these factors signals an underappreciation of their importance. On the contrary, socio-cultural aspects should be carefully considered by project developers, since the harmonization of health data cannot ignore the cultural peculiarities of the single contexts from where data are pooled [384]. Harmonization, linking and sharing do not happen in a vacuum and opening up the dialogue between data processors and society at large can be an important success factor for the harmonization of health data in the long run.

Limitations

The limitations of this systematic review include choices that we made regarding the number of databases used for our search, the fact that we did search using English key words, and that only 20 percent of included papers went through double checking for data extraction consistency. We could have thus missed valuable studies that were published only in Danish, French, and German which we could have found if key words had been in those languages. Given the high number of papers included and resources related constraints, we were unable to double check for all information recorded, but in light of low discrepancies found in the portion of records which were double-checked, we are confident in our output. A reporting bias of barriers and facilitators identified in the included papers cannot be excluded as published papers are focused mostly on the effectiveness of their interventions rather than on the implementation phase. It is possible that our results are thus biased towards barriers and facilitators more likely to be reported in the papers (e.g. those of a technical nature). Given the low numbers of certain types of reported barriers and facilitators, it is difficult to compare the situation in the two countries without under- or over-exaggerating their presence or absence in the two countries.

However, the main objective of this systematic review was to identify barriers and facilitators to harmonized health data collection, sharing and linkage in Denmark and Switzerland. Causal inference was not part of this review's primary objectives.

3.6 Conclusion

This systematic review gathered evidence from Switzerland and Denmark to map and describe barriers and facilitators concerning data harmonization, sharing and linkage. Given the focus of this review on Switzerland and Denmark, part of the findings has specific relevance for these two countries. In particular, for Switzerland it has emerged that fragmentation in the health data environment is a key challenge for harmonizing, sharing and linking of data. Since the implementation of more centralized governance systems—which are of great use in Denmark—might not be a viable option for Switzerland because of the political structure of the country, a distributed governance model, which emphasises interoperability of health data, seems to be the preferable way forward. The introduction of Blockchain technology for patient records, which insures security and respects decentralization [385, 386], is reportedly an auspicious technology as its use in the Estonian healthcare system described by Mettler [387] suggests. This review outlined that the existing data infrastructure at the national-level in Denmark incentivizes the completion of retrospective registry-based studies relying on data reuse. Although barriers are still reported, the existence and comprehensiveness of this data infrastructure confirms that past efforts to improve the health data framework have proven successful. For the future, efforts should focus on easing projects involving cross-national collaborations.

However, other findings are meaningful well beyond the borders of the two countries specifically considered. In particular, in this review it has emerged that, although a great number of barriers and facilitators are mentioned by the projects involving health data harmonization, sharing and linking, reporting focusses predominantly on specific aspects–above all technical ones. Whereas technical aspects are certainly important, the reluctancy to mention also issues of other natures is detrimental to the more general effort of the scientific community to favour the harmonization of health data. Referring more openly to the difficulties encountered at the ethico-legal level, for example, might be of help both for new projects to develop appropriate approaches and for policy makers to gather evidence on which regulatory interventions are needed. The under-appreciation of ethico-legal, socio-cultural and other context-specific complexities is a faux-pas, since the trust of both data-subjects and society at large is indispensable for the success a community in improving the health data context, like the

experience of Iceland has demonstrated in the past [388]. There, the project to build a national "health sector database" with health information of all citizens imported from their medical records failed also due to the underappreciation of ethico-legal issues (e.g. informed consent and privacy). Specifically, the population complained that inclusion of personal medical records into the database was supposed to happen without consent by individuals or the possibility to opt out. This was felt like a violation of privacy, because of the risk of re-identification and also due to the fact that the database was supposed to be run by a private company [389]. A privacy complaint was brought in front of the national high court, who ruled against the project to build the database. For this reason, the project was definitely aborted [390].

In summary, the success of current and future projects is likely to depend on a better understanding and appreciation of the complexities associated with harmonizing, sharing and linking health data. In the same line, proposed solutions to harmonization issues should not underestimate the contextual particularities of the country, in which such health data processes occur.

Acknowledgments

The authors acknowledge the financial support received from the Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356). The SNSF is not involved in any other aspect of this review and the views expressed are those of the authors and not necessarily those of the funder. This systematic review is part of a broader project funded by the SNSF titled "*Advancing Smart solutions and eliminating barriers for the Acquisition*, *Analysis and Sharing of Health data in Switzerland* (SMAASH)". MCM was an independent researcher when the work for this systematic review was carried out and she is currently affiliated with the Institute for Social and Preventive Medicine (ISPM) at the University of Bern.

Author contributions

Conceptualization: Tenzin Wangmo, Bernice Simone Elger.

Data curation: Lester Darryl Geneviève, Andrea Martani, Maria Christina Mallet.

Formal analysis: Lester Darryl Geneviève.

Funding acquisition: Bernice Simone Elger.

Investigation: Lester Darryl Geneviève, Andrea Martani, Maria Christina Mallet.

Methodology: Lester Darryl Geneviève, Andrea Martani, Tenzin Wangmo.

Project administration: Tenzin Wangmo, Bernice Simone Elger.

Resources: Bernice Simone Elger.

Supervision: Tenzin Wangmo, Bernice Simone Elger.

Validation: Tenzin Wangmo.

Visualization: Lester Darryl Geneviève.

Writing – original draft: Lester Darryl Geneviève, Andrea Martani, Tenzin Wangmo, Bernice Simone Elger.

Writing – review & editing: Lester Darryl Geneviève, Andrea Martani, Maria Christina Mallet, Tenzin Wangmo, Bernice Simone Elger

Competing Interests

I have read the journal's policy and the authors of this manuscript report the following non-financial competing interests: LDG and MCM are married and this does not alter our adherence to PLOS ONE policies on sharing data and materials.

Data availability statement

All relevant data are included within the paper and were obtained from articles available in databases quoted in the manuscript.

Funding

LDG, AM, TW and BSE acknowledge the financial support of the Swiss National Science Foundation (SNFNRP-74 Smarter Health Care, grant number 407440_167356). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. URL of funder: http://www.snf.ch/en/Pages/default.aspx.

3.7 References

1. Mostert M, Bredenoord AL, Biesaart MC, van Delden JJ. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet. 2016; 24 (7):1096.
2. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, et al. Making sense of big data in health research: Towards an EU action plan. Genome Med. 2016; 8(1).
3. de Carvalho EC, Batilana AP, Simkins J, Martins H, Shah J, Rajgor D, et al. Application description and policy model in collaborative environment for sharing of information on epidemiological and clinical research data sets. PloS one. 2010; 5(2):e9314. https://doi.org/10.1371/journal.pone.0009314 PMID: 20174560

4. Largent EA, Joffe S, Miller FG. Can research and care be ethically integrated? The Hastings Center report. 2011; 41(4):37–46. https://doi.org/10.1002/j.1552-146x.2011.tb00123.x PMID: 21845922

5. NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development. Data Harmonization 2019 [Internet] Available from: https://www.icpsr.umich.edu/icpsrweb/content/DSDR/ harmonization.html.

6. Institute of Medicine (US). Standardization to Enhance Data Sharing. Sharing Clinical Research Data: Workshop Summary: National Academies Press (US); 2013.

7. Blasimme A, Fadda M, Schneider M, Vayena E. Data Sharing For Precision Medicine: Policy Lessons And Future Directions. Health Aff (Millwood). 2018; 37(5):702–9.

8. OECD. New Health Technologies2017.

9. IOM (Institute of Medicine). Sharing Clinical Research Data: National Academies Press (US); 2013.

10. Chambers DA, Feero WG, Khoury MJ. Convergence of Implementation Science, Precision Medicine, and the Learning Health Care System: A New Model for Biomedical Research. JAMA. 2016; 315 (18):1941–2. https://doi.org/10.1001/jama.2016.3867 PMID: 27163980

11. Burton PR, Fortier I, Deschênes M, Hansell A, Palmer LJ. Biobanks and biobank harmonisation. An Introduction to Genetic Epidemiology. Bristol, UK: The Policy Press; 2011. p. 155–74.

12. Fortier I, Doiron D, Burton P, Raina P. Invited Commentary: Consolidating Data Harmonization—How to Obtain Quality and Applicability? Am J Epidemiol. 2011; 174(3):261–4. https://doi.org/10.1093/aje/ kwr194 PMID: 21749975

13. Zwitter A. Big Data ethics. Big Data & Society. 2014; 1(2):2053951714559253.

14. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. Sci Eng Ethics. 2016; 22(2):303–41. https://doi.org/10.1007/s11948-015-9652-2 PMID: 26002496

15. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014; 14:1144. https://doi.org/10.1186/ 1471-2458-14-1144 PMID: 25377061

16. Edwards A, Hollin I, Barry J, Kachnowski S. Barriers to cross—institutional health information exchange: a literature review. JHIM. 2010; 24(3):22–34. PMID: 20677469

17. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc. 2010; 17(3):288–94. https://doi.org/10.1136/jamia.2010.003673 PMID: 20442146

18. Kierkegaard P. eHealth in Denmark: a case study. J Med Syst. 2013; 37(6):9991. https://doi.org/10. 1007/s10916-013-9991-y PMID: 24166019

19. Hoeyer K. Denmark at a Crossroad? Intensified Data Sourcing in a Research Radical Country. In: Mittelstadt BD, Floridi L, editors. The Ethics of Biomedical Big Data. Cham: Springer International Publishing; 2016. p. 73–93.

20. Meier-Abt PJ, Lawrence AK, Selter L, Vayena E, Schwede T. The Swiss approach to precision medicine. Swiss Med Wkly. 2018.

21. Wallace LS. A view of health care around the world. Ann Fam Med. 2013; 11(1):84.

22. De Pietro C, Francetic I. E-health in Switzerland: The laborious adoption of the federal law on electronic health records (EHR) and health information exchange (HIE) networks. Health Policy. 2017.

23. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and metaanalyses: the PRISMA statement. J Clin Epidemiol. 2009; 62(10):1006–12. https://doi.org/10.1016/j.jclinepi.2009.06.005 PMID: 19631508

24. Office Fédéral de la Santé Publique (OFSP). Stratégie eHealth Suisse 2017 [Internet] Available from: https://www.bag.admin.ch/bag/fr/home/themen/strategien-politik/nationale-gesundheitsstrategien/ strategie-ehealth-schweiz.html.

25. Schmidt B-M, Colvin CJ, Hohlfeld A, Leon N. Defining and conceptualising data harmonisation: a scoping review protocol. Syst Rev. 2018; 7(1):226. https://doi.org/10.1186/s13643-018-0890-7 PMID: 30522527

26. Vayena E, Dzenowagis J, Brownstein JS, Sheikh A. Policy implications of big data in the health sector. Bull World Health Organ. 2018; 96(1):66–8. https://doi.org/10.2471/BLT.17.197426 PMID: 29403102

27. Popay JR, H. M.; Sowden A.; Petticrew M.; Arai L.; Rodgers M.; Britten N. Guidance on the conduct of narrative synthesis in sytematic reviews. 2006.

28. Aabakke AJM, Krebs L, Ladelund S, Secher NJ. Incidence of Incisional Hernia after Cesarean Delivery: A Register-Based Cohort Study. PloS one. 2014; 9(9).

29. Adam M, von der Weid N, Michel G, Zwahlen M, Lutz J-M, Probst-Hensch N, et al. Access to Specialized Pediatric Cancer Care in Switzerland. Pediatr Blood Cancer. 2010; 54(5):721–7. https://doi.org/10.1002/pbc.22426. PMID: 20108340

30. Agergaard K, Mau-Sørensen M, Stage TB, Jørgensen TL, Hassel RE, Steffensen KD, et al. Clopidogrel–Paclitaxel Drug–Drug Interaction: A Pharmacoepidemiologic Study. Clin Pharmacol Ther. 2017;102(3):547–53. https://doi.org/10.1002/cpt.674 PMID: 28224612

31. Agten A, Passweg D, von Orelli S, Ringel N, Tschudi R, Tutschek B. Temporal trends of postpartum haemorrhage in Switzerland: a 22-year retrospective population-based cohort study. Swiss Med Wkly. 2017; 147.

32. Ammundsen HB, Faber MT, Jensen A, Hogdall E, Blaakaer J, Hogdall C, et al. Use of analgesic drugs and risk of ovarian cancer: results from a Danish case-control study. Acta Obst Gynecol Scand. 2012; 91(9):1094–102.

33. Andersen JS, Olivarius Nde F, Krasnik A. The Danish National Health Service Register. Scand J Public Health. 2011; 39(7 Suppl):34–7. https://doi.org/10.1177/1403494810394718 PMID: 21775348

34. Andersen P, Andersen LM, Iversen LH. Iatrogenic ureteral injury in colorectal cancer surgery: a nationwide study comparing laparoscopic and open approaches. Surg Endosc. 2014.

35. Andres M, Feller A, Arndt V, Grp NW. Trends of incidence, mortality, and survival of multiple myeloma in Switzerland between 1994 and 2013. Cancer Epidemiol. 2018; 53:105–10. https://doi.org/10.1016/j. canep.2018.01.015 PMID: 29414629

36. Annaheim B, Wangmo T, Bretschneider W, Vogel M, Elger BS. Can routine data from prisoners' files be used to estimate prevalence rates of illicit drug use among prisoners? Int J Public Health. 2018; 63(1):33–40. https://doi.org/10.1007/s00038-017-1030-1. PMID: 28835980

37. Antonsen SL, Meyhoff CS, Lundvall L, Haøgdall C. Surgical-site infections and postoperative complications: Agreement between the Danish Gynecological Cancer Database and a randomized clinical trial. Acta Obstet Gynecol Scand. 2011; 90(1):72–6. https://doi.org/10.1111/j.1600-0412.2010.01007. x. PMID: 21275918

38. Antonsen K, Rosenstock CV, Lundstrom LH. The Danish Anaesthesia Database. Clin Epidemiol. 2016; 8:435–8. https://doi.org/10.2147/CLEP.S99517 PMID: 27843340

39. Arboe B, Josefsson P, Jørgensen J, Haaber J, Jensen P, Poulsen C, et al. Danish National Lymphoma Registry. Clin Epidemiol. 2016; 8:577–81. https://doi.org/10.2147/CLEP.S99470. PMID: 27822102

40. Arking DE, Pulit SL, Crotti L, Van Der Harst P, Munroe PB, Koopmann TT, et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. Nat Genet. 2014; 46(8):826–36. https://doi.org/10.1038/ng.3014 PMID: 24952745

41. Atlado´ttir HO´, Henriksen TB, Schendel DE, Parner ET. Using maternally reported data to investigate the association between early childhood infection and autism spectrum disorder: The importance of data source. Paediatr Perinat Epidemiol. 2012; 26(4):373–85. https://doi.org/10.1111/j.1365-3016. 2012.01296.x. PMID: 22686389

42. Aubert CE, Streit S, Da Costa BR, Collet T-H, Cornuz J, Gaspoz J-M, et al. Polypharmacy and specific comorbidities in university primary care settings. Eur J Intern Med. 2016; 35:35–42. https://doi.org/10. 1016/j.ejim.2016.05.022. PMID: 27289492

43. Auer R, Gencer B, Raeber L, Klingenberg R, Carballo S, Carballo D, et al. Quality of Care after Acute Coronary Syndromes in a Prospective Cohort with Reasons for Non-Prescription of Recommended Medications. PloS one. 2014; 9(3).

44. Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour JC, et al. Harmonization process for the identification of medical events in eight European healthcare databases: The experience from the EU-ADR project. J Am Med Inform Assoc. 2013; 20(1):184–92. https://doi.org/10.1136/amiajnl-2012000933 PMID: 22955495

45. Avlund TH, Erichsen R, Iversen LH. Sensitivity and positive predictive value of the registration of selfexpanding metal stent treatment for obstructive colorectal cancer in two danish nationwide registries. Clin Epidemiol. 2018; 10:1411–5. https://doi.org/10.2147/CLEP.S174867 PMID: 30349389

46. Bachelet D, Hassler S, Mbogning C, Link J, Ryner M, Ramanujam R, et al. Occurrence of Anti-Drug Antibodies against Interferon-Beta and Natalizumab in Multiple Sclerosis: A Collaborative Cohort Analysis. PloS one. 2016; 11(11).

47. Baker JL, Olsen LW, Andersen I, Pearson S, Hansen B, Sørensen TI. Cohort profile: the Copenhagen School Health Records Register. Int J Epidemiol. 2009; 38(3):656–62. https://doi.org/10.1093/ije/ dyn164 PMID: 18719090

48. Baldur-Felskov B, Kjaer SK, Albieri V, Steding-Jessen M, Kjaer T, Johansen C, et al. Psychiatric disorders in women with fertility problems: results from a large Danish register-based cohort study. Hum Reprod. 2013; 28(3):683–90. https://doi.org/10.1093/humrep/des422 PMID: 23223399

49. Balgobind BV, Raimondi SC, Harbott J, Zimmermann M, Alonzo TA, Auvrignon A, et al. Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: Results of an international retrospective study. Blood. 2009; 114(12):2489–96. https://doi.org/10.1182/blood-2009-04215152. PMID: 19528532

50. Bay-Nielsen M, Kehlet H. Anaesthesia and post-operative morbidity after elective groin hernia repair: a nation-wide study. Acta Anaesthesiol Scand. 2008; 52(2):169–74. https://doi.org/10.1111/j.13996576.2007.01514.x. PMID: 17999709

51. Beduneau G, Pham T, Schortgen F, Piquilloud L, Zogheib E, Jonas M, et al. Epidemiology of Weaning Outcome according to a New Definition The WIND Study. Am J Respir Crit Care Med. 2017; 195(6):772–83. https://doi.org/10.1164/rccm.201602-0320OC. PMID: 27626706

52. Begre S, Traber M, Gerber M, von Kaenel R. Regional origin and decrease of pain in patients with depressive symptoms under treatment with venlafaxine. Soc Psychiatry Psychiatr Epidemiol. 2010; 45 (1):17–24. https://doi.org/10.1007/s00127-009-0036-2. PMID: 19300890

53. Bendixen M, Kronborg C, Jorgensen OD, Andersen C, Licht PB. Cost-utility analysis of minimally invasive surgery for lung cancer: a randomized controlled trial. Eur J Cardio-thorac Surg. 2019.

54. Beretta-Piccoli BT, Stirnimann G, Cerny A, Semela D, Hessler R, Helbling B, et al. Geoepidemiology of Primary Biliary Cholangitis: Lessons from Switzerland. Clin Rev Allergy Immunology. 2017.

55. Binderup MLM, Stendell AS, Galanakis M, Møller HU, Kiilgaard JF, Bisgaard ML. Retinal hemangioblastoma: Prevalence, incidence and frequency of underlying von Hippel-Lindau disease. Br J Ophthalmol. 2018; 102(7):942–7. https://doi.org/10.1136/bjophthalmol-2017-310884. PMID: 28972023

56. Bisgaard H, Vissing NH, Carson CG, Bischoff AL, Folsgaard NV, Kreiner-Moller E, et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. Clin Exp Allergy. 2013; 43(12):1384–94. https://doi.org/10.1111/cea.12213. PMID: 24118234

57. Bjerregaard B, Larsen OB. The Danish Pathology Register. Scand J Public Health. 2011; 39(7 Suppl):72–4. https://doi.org/10.1177/1403494810393563. PMID: 21775357

58. Bjornholt SM, Kjaer SK, Nielsen TSS, Jensen A. Risk for borderline ovarian tumours after exposure to fertility drugs: results of a population-based cohort study. Hum Reprod. 2015; 30(1):222–31. https:// doi.org/10.1093/humrep/deu297 PMID: 25376453

59. Blaha J, Barteczko-Grajek B, Berezowicz P, Charvat J, Chvojka J, Grau T, et al. Space GlucoseControl system for blood glucose control in intensive care patients—a European multicentre observational study. BMC Anesthesiol. 2016; 16(1).

60. Blenstrup LT, Knudsen LB. Danish registers on aspects of reproduction. Scand J Public Health. 2011;39(7 Suppl):79–82. https://doi.org/10.1177/1403494811399957. PMID: 21775359

61. Blichert-Toft M, Christiansen P, Mouridsen HT. Danish Breast Cancer Cooperative Group—DBCG: History, organization, and status of scientific achievements at 30-year anniversary. Acta Oncol. 2008;47(4):497–505. https://doi.org/10.1080/02841860802068615. PMID: 18465316

62. Bodin CR, Rasmussen MM, Tabor A, Westbom L, Tiblad E, Ekelund CK, et al. Ultrasound in Prenatal Diagnostics and Its Impact on the Epidemiology of Spina Bifida in a National Cohort from Denmark with a Comparison to Sweden. Biomed Res Int. 2018.

63. Boje CR, Dalton SO, Primdahl H, Kristensen CA, Andersen E, Johansen J, et al. Evaluation of comorbidity in 9388 head and neck cancer patients: A national cohort study from the DAHANCA database. Radiother Oncol. 2014; 110(1):91–7. https://doi.org/10.1016/j.radonc.2013.11.009 PMID: 24412015

64. Brenner R, Waeber B, Allemann Y. Medical treatment of hypertension in Switzerland. The 2009 Swiss Hypertension Survey (SWISSHYPE). Swiss Med Wkly. 2011; 141:w13169. https://doi.org/10.4414/ smw.2011.13169 PMID: 21374529

65. Brink C, Lorenzen EL, Krogh SL, Westberg J, Berg M, Jensen I, et al. DBCG hypo trial validation of radiotherapy parameters from a national data bank versus manual reporting. Acta Oncol. 2018; 57(1):107–12. https://doi.org/10.1080/0284186X.2017.1406140. PMID: 29202666

66. Burgstaller JM, Held U, Brunner F, Porchet F, Farshad M, Steurer J, et al. The Impact of Obesity on the Outcome of Decompression Surgery in Degenerative Lumbar Spinal Canal Stenosis: Analysis of the Lumbar Spinal Outcome Study (LSOS): A Swiss Prospective, Multicenter Cohort Study. Spine. 2016; 41(1):82–9. https://doi.org/10.1097/BRS.0000000000001128. PMID: 26689396

67. Cainzos-Achirica M, Varas-Lorenzo C, Pottegard A, Asmar J, Plana E, Rasmussen L, et al. Methodological challenges when evaluating potential off-label prescribing of drugs using electronic health care databases: A case study of dabigatran etexilate in Europe. Pharmacoepidemio Drug Saf. 2018; 27 (7):713–23.

68. Calhaz-Jorge C, De Geyter C, Kupka MS, De Mouzon J, Erb K, Mocanu E, et al. Assisted reproductive technology in Europe, 2013: Results generated from European registers by ESHRE. Hum Reprod. 2017; 32(10):1957–73. https://doi.org/10.1093/humrep/dex264. PMID: 29117383

69. Calvet D, Mas JL, Algra A, Becquemin JP, Bonati LH, Dobson J, et al. Carotid stenting: Is there an operator effect? A pooled analysis from the carotid stenting trialists' collaboration. Stroke. 2014; 45(2):527–32. https://doi.org/10.1161/STROKEAHA.113.003526. PMID: 24347422

70. Carstensen B, Kristensen JK, Ottosen P, Borch-Johnsen K. The Danish National Diabetes Register: trends in incidence, prevalence and mortality. Diabetologia. 2008; 51(12):2187–96. https://doi.org/10.1007/s00125-008-1156-z. PMID: 18815769

71. Caspersen S, Elkjaer M, Riis L, Pedersen N, Mortensen C, Jess T, et al. Infliximab for Inflammatory Bowel Disease in Denmark 1999–2005: Clinical Outcome and Follow-Up Evaluation of Malignancy and Mortality. Clin Gastroenterol Hepatol. 2008; 6(11):1212–7. https://doi.org/10.1016/j.cgh.2008.05. 010. PMID: 18848503

72. Chaigne B, Chizzolini C, Perneger T, Trendelenburg M, Huynh-Do U, Dayer E, et al. Impact of disease activity on health-related quality of life in systemic lupus erythematosus—a cross-sectional analysis of the Swiss Systemic Lupus Erythematosus Cohort Study (SSCS). BMC Immunol. 2017; 18(1):17. https://doi.org/10.1186/s12865-017-0200-5. PMID: 28351341

73. Chesnaye N, Bonthuis M, Schaefer F, Groothoff JW, Verrina E, Heaf JG, et al. Demographics of paediatric renal replacement therapy in Europe: a report of the ESPN/ERA–EDTA registry. Pediatr Nephrol. 2014; 29(12):2403–10. https://doi.org/10.1007/s00467-014-2884-6 PMID: 25039018

74. Chmiel C, Bhend H, Senn O, Zoller M, Rosemann T. The FIRE project: a milestone for research in primary care in Switzerland. Swiss Med Wkly. 2011; 140:w13142. https://doi.org/10.4414/smw.2011. 13142 PMID: 21279858

75. Christensen BO, Overgaard J, Kettner LO, Damsgaard TE. Long-term evaluation of postmastectomy breast reconstruction. Acta Oncol. 2011; 50(7):1053–61. https://doi.org/10.3109/0284186X.2011. 584554 PMID: 21745130

76. Christensen AS, Roed C, Omland LH, Andersen PH, Obel N, Andersen AB. Long-term mortality in patients with tuberculous meningitis: a Danish nationwide cohort study. PloS one. 2011; 6(11): e27900. https://doi.org/10.1371/journal.pone.0027900 PMID: 22132165

77. Christensen S, Johansen MB, Tonnesen E, Larsson A, Pedersen L, Lemeshow S, et al. Preadmission beta-blocker use and 30-day mortality among patients in intensive care: a cohort study. Critical Care. 2011;15(2).

78. Christensen AS, Roed C, Andersen PH, Andersen AB, Obel N. Long-term mortality in patients with pulmonary and extrapulmonary tuberculosis: a Danish nationwide cohort study. Clin Epidemiol. 2014;6:405–21. https://doi.org/10.2147/CLEP.S65331. PMID: 25419160

79. Christensen DH, Rungby J, Thomsen RW. Nationwide trends in glucose-lowering drug use, Denmark, 1999–2014. Clin Epidemiol. 2016; 8:381–7. https://doi.org/10.2147/CLEP.S113211. PMID: 27789974

80. Christensen TD, Skjoth F, Nielsen PB, Maegaard M, Grove EL, Larsen TB. Self-Management of Anticoagulant Therapy in Mechanical Heart Valve Patients: A Matched Cohort Study. Ann Thorac Surg.2016; 101(4):1494–9. https://doi.org/10.1016/j.athoracsur.2015.09.084. PMID: 26572254

81. Christiansen P, Bjerre K, Al-Suliman N, Moller S, Dbcg. Recurrence pattern and prognosis in low-risk breast cancer patients—Data from the DBCG 89-A programme. Acta Oncologica. 2008; 47(4):691–703. https://doi.org/10.1080/02841860802056594. PMID: 18465337

82. Christiansen P, Friis E, Balslev E, Jensen D, Moller S. Sentinel node biopsy in breast cancer: five years experience from Denmark. Acta Oncol. 2008; 47(4):561–8. https://doi.org/10.1080/02841860802023206. PMID: 18465323

83. Christoffersen MW, Brandt E, Oehlenschlager J, Rosenberg J, Helgstrand F, Jorgensen LN, et al. No difference in incidence of port-site hernia and chronic pain after single-incision laparoscopic cholecystectomy versus conventional laparoscopic cholecystectomy: a nationwide prospective, matched cohort study. Surgical Endoscopy and Other Interventional Techniques. 2015; 29(11):3239–45. https://doi.org/10.1007/s00464-015-4066-4. PMID: 25612547

84. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. Lancet. 2011; 377(9760):127–38. https://doi.org/10.1016/S0140-6736(10)62231-3. PMID: 21183212

85. Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: The EU-ADR Project. Pharmacoepidemiol Drug Saf. 2011; 20(1):1–11. https://doi.org/10.1002/pds.2053. PMID: 21182150

86. Corraini P, Heide-Jorgensen U, Schiodt M, Norholt SE, Acquavella J, Sorensen HT, et al. Osteonecrosis of the jaw and survival of patients with cancer: a nationwide cohort study in Denmark. Cancer Med. 2017; 6(10):2271–7. https://doi.org/10.1002/cam4.1173. PMID: 28941210

87. Costantino G, Ruwald MH, Quinn J, Camargo CA Jr., Dalgaard F, Gislason G, et al. Prevalence of Pulmonary Embolism in Patients With Syncope. JAMA Intern Med. 2018; 178(3):356–62. https://doi.org/10.1001/jamainternmed.2017.8175. PMID: 29379959

88. Cotter AG, Vrouenraets SME, Brady JJ, Wit FW, Fux CA, Furrer H, et al. Impact of switching from zidovudine to tenofovir disoproxil fumarate on bone mineral density and markers of bone metabolism in virologically suppressed HIV-1 infected patients; A substudy of the PREPARE study. J Clin Endocrinol Metab. 2013; 98(4):1659–66. https://doi.org/10.1210/jc.2012-3686. PMID: 23436922

89. Czauderna P, Haeberle B, Hiyama E, Rangaswami A, Krailo M, Maibach R, et al. The Children's Hepatic tumors International Collaboration (CHIC): Novel global rare tumor database yields new prognostic factors in hepatoblastoma and becomes a research model. Eur J Cancer. 2016; 52:92–101. https://doi.org/10.1016/j.ejca.2015.09.023. PMID: 26655560

90. Dalgard O, Bjoro K, Ring-Larsen H, Verbaan H. In patients with HCV genotype 2 or 3 infection and RVR 14 weeks treatment is noninferior to 24 weeks. Pooled analysis of two Scandinavian trials. Eur J Gastroenterol Hepatol. 2010; 22(5):552–6. https://doi.org/10.1097/MEG.0b013e328335b29e. PMID: 20154627

91. Damgaard OE, Jensen M-B, Kroman N, Tvedskov TF. Quantifying the number of lymph nodes identified in one-stage versus two-stage axillary dissection in breast cancer. Breast. 2013; 22(1):44–6. https://doi.org/10.1016/j.breast.2012.03.009. PMID: 22494665

92. Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Bronnum D, et al. Risk of Ischemic Heart Disease in Women after Radiotherapy for Breast Cancer. N Engl J Med. 2013; 368(11):987–98. https://doi.org/10.1056/NEJMoa1209825. PMID: 23484825

93. Dastani Z, Hivert MF, Timpson N, Perry JRB, Yuan X, Scott RA, et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: A multi-ethnic meta-analysis of 45,891 individuals. PLoS Genet. 2012; 8(3).

94. De Angelis R, Francisci S, Baili P, Marchesi F, Roazzi P, Belot A, et al. The EUROCARE-4 database on cancer survival in Europe: Data standardisation, quality control and methods of statistical analysis. Eur J Cancer. 2009; 45(6):909–30. https://doi.org/10.1016/j.ejca.2008.11.003. PMID: 19128955

95. De Groot MCH, Schuerch M, De Vries F, Hesse U, Oliva B, Gil M, et al. Antiepileptic drug use in seven electronic health record databases in Europe: A methodologic comparison. Epilepsia. 2014; 55(5):666–73. https://doi.org/10.1111/epi.12557. PMID: 24575970

96. della Torre R, Combescure C, Cortes B, Marazza G, Beltraminelli H, Naldi L, et al. Clinical presentation and diagnostic delay in bullous pemphigoid: a prospective nationwide cohort. Br J Dermatol. 2012;167(5):1111–7. https://doi.org/10.1111/j.1365-2133.2012.11108.x. PMID: 22709136

97. Dencker D, Pedersen F, Engstrom T, Kober L, Hojberg S, Nielsen MB, et al. Major femoral vascular access complications after coronary diagnostic and interventional procedures: A Danish register study. Int J Cardiol. 2016; 202:604–8. https://doi.org/10.1016/j.ijcard.2015.09.018. PMID: 26447671

98. De Vos Andersen NB, Kent P, Hjort J, Christiansen DH. Clinical course and prognosis of musculoskeletal pain in patients referred for physiotherapy: does pain site matter? BMC Musculoskelet Disord. 2017; 18(1).

99. Diel P, Reuss W, Aghayev E, Moulin P, Ro¨der C. SWISSspine—A nationwide health technology assessment registry for balloon kyphoplasty: Methodology and first results. Spine. 2010; 10(11):961– 71.

100. Disanto G, Benkert P, Lorscheider J, Mueller S, Vehoff J, Zecca C, et al. The Swiss Multiple Sclerosis Cohort-Study (SMSC): A Prospective Swiss Wide Investigation of Key Phases in Disease Evolution and New Treatment Options. PloS one. 2016; 11(3).

101. Donia M, Kimper-Karl ML, Hoyer KL, Bastholt L, Schmidt H, Svane IM. The majority of patients with metastatic melanoma are not represented in pivotal phase III immunotherapy trials. Eur J Cancer. 2017; 74:89–95. https://doi.org/10.1016/j.ejca.2016.12.017 PMID: 28335891

102. Downs J, Stahlhut M, Wong K, Syhler B, Bisgaard A-M, Jacoby P, et al. Validating the Rett Syndrome Gross Motor Scale. PloS one. 2016; 11(1).

103. Dreyer NA, Blackburn S, Hliva V, Mt-Isa S, Richardson J, Jamry-Dziurla A, et al. Balancing the Interests of Patient Data Protection and Medication Safety Monitoring in a Public-Private Partnership. JMIR Med Inform. 2015; 3(2).

104. Edgren G, Rostgaard K, Vasan SK, Wikman A, Norda R, Pedersen OB, et al. The new Scandinavian Donations and Transfusions database (SCANDAT2): a blood safety resource with added versatility. Transfusion. 2015; 55(7):1600–6. https://doi.org/10.1111/trf.12986. PMID: 25573303

105. Ehlers L, Overvad K, Sorensen J, Christensen S, Bech M, Kjolby M. Analysis of cost effectiveness of screening Danish men aged 65 for abdominal aortic aneurysm. BMJ. 2009; 338.

106. Ekelund CK, Petersen OB, Jorgensen FS, Kjaergaard S, Larsen T, Olesen AW, et al. The Danish Fetal Medicine Database: establishment, organization and quality assessment of the first trimester screening program for trisomy 21 in Denmark 2008–2012. Acta Obstet Gynecol Scand. 2015; 94(6):577–83. https://doi.org/10.1111/aogs.12581. PMID: 25597330

107. El-Galaly TC, Jakobsen LH, Hutchings M, de Nully Brown P, Nilsson-Ehle H, Szekely E, et al. Routine Imaging for Diffuse Large B-Cell Lymphoma in First Complete Remission Does Not Improve PostTreatment Survival: A Danish-Swedish Population-Based Study. J Clin Oncol. 2015; 33(34):3993–8. https://doi.org/10.1200/JCO.2015.62.0229. PMID: 26438115

108. Elliott A, Mork TJ, Hojlund M, Christensen T, Jeppesen R, Madsen N, et al. QTc interval in patients with schizophrenia receiving antipsychotic treatment as monotherapy or polypharmacy. CNS Spectr. 2017:1–6.

109. Engelberger RP, Noll G, Schmidt D, Alatri A, Frei B, Kaiser WE, et al. Initiation of rivaroxaban in patients with nonvalvular atrial fibrillation at the primary care level: The Swiss Therapy in Atrial Fibrillation for the Regulation of Coagulation (STAR) Study. Eur J Intern Med. 2015; 26(7):508–14. https://doi.org/10.1016/j.ejim.2015.04.014. PMID: 25935131

110. Erdem H, Ozturk-Engin D, Tireli H, Kilicoglu G, Defres S, Gulsun S, et al. Hamsi scoring in the prediction of unfavorable outcomes from tuberculous meningitis: results of Haydarpasa-II study. J Neurol. 2015; 262(4):890–8. https://doi.org/10.1007/s00415-015-7651-5. PMID: 25634680

111. Erichsen R, Lash TL, Hamilton-Dutoit SJ, Bjerregaard B, Vyberg M, Pedersen L. Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank. Clin Epidemiol.2010; 2:51–6. https://doi.org/10.2147/clep.s9908. PMID: 20865103

112. Erichsen R, Christiansen CF, Froslev T, Jacobsen J, Sorensen HT. Intravenous bisphosphonate therapy and atrial fibrillation/flutter risk in cancer patients: a nationwide cohort study. Br J Cancer. 2011;105(7):881–3. https://doi.org/10.1038/bjc.2011.338. PMID: 21878939

113. Erlangsen A, Canudas-Romo V, Conwell Y. Increased use of antidepressants and decreasing suicide rates: a population-based study using Danish register data. J Epidemiol Community Health. 2008; 62(5):448–54. https://doi.org/10.1136/jech.2007.061580. PMID: 18413459

114. Escala-Garcia M, Guo Q, Do¨rk T, Canisius S, Keeman R, Dennis J, et al. Genome-wide association study of germline variants and breast cancer-specific mortality. Br J Cancer. 2019; 120(6):647–57. https://doi.org/10.1038/s41416-019-0393-x. PMID: 30787463

115. Escott-Price V, Bellenguez C, Wang LS, Choi SH, Harold D, Jones L, et al. Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. PloS one. 2014; 9(6).

116. Fago¨-Olsen CL, Ottesen B, Kehlet H, Markauskas A, Mosgaard BJ, Ottosen C, et al. Neoadjuvant chemotherapy as ovarian cancer treatment: ever more used with major regional differences. Dan Med J. 2012; 59(8).

117. Fahrner R, Malinka T, Klasen J, Candinas D, Beldi G. Additional surgical procedure is a risk factor for surgical site infections after laparoscopic cholecystectomy. Langenbeck Arch Surg. 2014; 399 (5):595–9.

118. Fedder J, Loft A, Parner ET, Rasmussen S, Pinborg A. Neonatal outcome and congenital malformations in children born after ICSI with testicular or epididymal sperm: a controlled national cohort study. Hum Reprod. 2013; 28(1):230–40. https://doi.org/10.1093/humrep/des377. PMID: 23154066

119. Fenger AQ, Helvind NM, Pommergaard H-C, Burcharth J, Rosenberg J. Fibrin sealant for mesh fixation in laparoscopic groin hernia repair does not increase long-term recurrence. Surgical Endoscopy and Other Interventional Techniques. 2016; 30(3):986–92. https://doi.org/10.1007/s00464-015-42800. PMID: 26092012

120. Fieten KB, Schappin R, Zijlstra WT, Figee L, Beutler J, Raymakers F, et al. Effectiveness of alpine climate treatment for children with difficult to treat atopic dermatitis: Results of a pragmatic randomized controlled trial (DAVOS trial). Clin Exp Allergy. 2018; 48(2):186–95. https://doi.org/10.1111/cea.13058. PMID: 29121432

121. Fløe A, Hilberg O, Wejse C, Ibsen R, Lokke A. Comorbidities, mortality and causes of death among patients with tuberculosis in Denmark 1998–2010: a nationwide, register-based case-control study. Thorax. 2018; 73(1):70–7. https://doi.org/10.1136/thoraxjnl-2016-209240. PMID: 28778918

122. Frandsen R, Baandrup L, Kjellberg J, Ibsen R, Jennum P. Increased all-cause mortality with psychotropic medication in Parkinson's disease and controls: a national register-based study. Parkinsonism Relat Disord. 2014; 20(11):1124–8. https://doi.org/10.1016/j.parkreldis.2014.07.012. PMID: 25164488

123. Frary EC, Gad D, Bastholt L, Hess S. The role of FDG-PET/CT in preoperative staging of sentinel lymph node biopsy-positive melanoma patients. EJNMMI Res. 2016; 6.

124. Freiberg FJ, Brynskov T, Munk MR, Sørensen TL, Wolf S, Wirth MA, et al. Low endophthalmitis rates after intravitreal anti-vascular endothelial grow factor injections in an operating room: A Retrospective Multicenter Study. Retina. 2017.

125. Friis S, Poulsen AH, Sorensen HT, Tjonneland A, Overvad K, Vogel U, et al. Aspirin and other non-steroidal anti-inflammatory drugs and risk of colorectal cancer: A Danish cohort study. Cancer Causes Control. 2009; 20(5):731–40. https://doi.org/10.1007/s10552-008-9286-7. PMID: 19122977

126. Funcke S, Sander M, Goepfert MS, Groesdonk H, Heringlake M, Hirsch J, et al. Practice of hemodynamic monitoring and management in German, Austrian, and Swiss intensive care units: the multicenter cross-sectional ICU-CardioMan Study. Ann Intensive Care. 2016; 6(1).

127. Furtwa¨ngler R, Kager L, Melchior P, Ru¨be C, Ebinger M, Nourkami-Tutdibi N, et al. High-dose treatment for malignant rhabdoid tumor of the kidney: No evidence for improved survival-The Gesellschaft fu¨r Pa¨diatrische Onkologie und Ha¨matologie (GPOH) experience. Pediatr Blood Cancer. 2018; 65(1): n/a-1.

128. Gammelager H, Christiansen CF, Johansen MB, Borre M, Schoonen M, Sorensen HT. Quality of urological cancer diagnoses in the Danish National Registry of Patients. Eur J Cancer Prev. 2012; 21(6):545–51. https://doi.org/10.1097/CEJ.0b013e328351c680. PMID: 22433631

129. Garcia-Etienne CA, Mansel RE, Tomatis M, Hei J, Biganzoli L, Ferrari A, et al. Trends in axillary lymph node dissection for early-stage breast cancer in Europe: Impact of evidence on practice. Breast. 2019; 45:89–96. https://doi.org/10.1016/j.breast.2019.03.002. PMID: 30925382

130. Gatta G, Trama A, Capocaccia R, Group RAW. Epidemiology of rare cancers and inequalities in oncologic outcomes. Eur J Surg Oncol. 2017.

131. Gatzioufas Z, Raiskup F, O'Brart D, Spoerl E, Panos GD, Hafezi F. Transepithelial corneal cross-linking using an enhanced riboflavin solution. J Refract Surg. 2016; 32(6):372–7. https://doi.org/10.3928/1081597X-20160428-02. PMID: 27304600

132. Geissbuhler A. Lessons learned implementing a regional health information exchange in Geneva as a pilot for the Swiss national eHealth strategy. Int J Med Inform. 2013; 82(5):e118–e24. https://doi.org/10.1016/j.ijmedinf.2012.11.002. PMID: 23332387

133. Ghith N, Ammari ABH, Rasmussen M, Frølich A, Cooper K, Tønnesen H. Impact of compliance on quit rates in a smoking cessation intervention: population study in Denmark. Clin Health Promot. 2012; 2:111–19.

134. Gjerstorff ML. The Danish Cancer Registry. Scand J Public Health. 2011; 39(7 Suppl):42–5. https:// doi.org/10.1177/1403494810393562. PMID: 21775350

135. Glintborg B, Ostergaard M, Dreyer L, Krogh NS, Tarp U, Hansen MS, et al. Treatment Response, Drug Survival, and Predictors Thereof in 764 Patients With Psoriatic Arthritis Treated With Anti-Tumor Necrosis Factor alpha Therapy Results From the Nationwide Danish DANBIO Registry. Arthritis Rheum. 2011; 63(2):382–90. https://doi.org/10.1002/art.30117. PMID: 21279995

136. Godballe C, Madsen AR, Pedersen HB, Sorensen CH, Pedersen U, Frisch T, et al. Post-thyroidectomy hemorrhage: a national study of patients treated at the Danish departments of ENT Head and Neck Surgery. Eur Arch Otorhinolaryngol. 2009; 266(12):1945–52. https://doi.org/10.1007/s00405009-0949-0. PMID: 19301027

137. Gorski M, Tin A, Garnaas M, McMahon GM, Chu AY, Tayo BO, et al. Genome-wide association study of kidney function decline in individuals of European descent. Kidney Int. 2015; 87(5):1017–29. https:// doi.org/10.1038/ki.2014.361. PMID: 25493955

138. Goutaki M, Maurer E, Halbeisen FS, Amirav I, Barbato A, Behan L, et al. The international primary ciliary dyskinesia cohort (iPCD Cohort): methods and first results. Eur Respir J. 2017; 49(1).

139. Goutaki M, Eich O, Halbeisen F, Barben J, Casaulta C, Clarenbach C, et al. The Swiss Primary Ciliary Dyskinesia registry: objectives, methods and first results. Swiss Med Wkly. 2019; 149.

140. Gradel KO, Dethlefsen C, Schønheyder HC, Nielsen H. Magnitude of bacteraemia is associated with increased mortality in non-typhoid salmonellosis: A one-year follow-up study. APMIS. 2008; 116(2):147–53. https://doi.org/10.1111/j.1600-0463.2008.00886.x. PMID: 18321366

141. Grann AF, Erichsen R, Nielsen AG, Frøslev T, Thomsen RW. Existing data sources for clinical epidemiology: The clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. Clin Epidemiol. 2011; 3:133–8. https://doi.org/10.2147/CLEP.S17901. PMID: 21487452

142. Gratwohl A, Pasquini MC, Aljurf M, Atsuta Y, Baldomero H, Foeken L, et al. One million haemopoietic stem-cell transplants: a retrospective observational study. Lancet Haematol. 2015; 2(3):e91–e100. https://doi.org/10.1016/S2352-3026(15)00028-9. PMID: 26687803

143. Gregersen R, Andresen K, Burcharth J, Pommergaard HC, Rosenberg J. Short-term mortality, readmission, and recurrence in treatment of acute diverticulitis with abscess formation: a nationwide register-based cohort study. Int J Colorectal Dis. 2016; 31(5):983–90. https://doi.org/10.1007/s00384-0162568-8 PMID: 27029799

144. Griffin SJ, Borch-Johnsen K, Davies MJ, Khunti K, Rutten GE, Sandbaek A, et al. Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): a cluster-randomised trial. Lancet. 2011; 378(9786):156–67. https://doi.org/10.1016/S0140-6736(11)60698-3. PMID: 21705063

145. Gromov K, Brix M, Kallemose T, Troelsen A. Early results and future challenges of the Danish Fracture Database. Dan Med J. 2014; 61(6).

146. Gruber P, Zeller S, Garcia-Esperon C, Berberat J, Anon J, Diepers M, et al. Embolus Retriever with Interlinked Cages versus other stent retrievers in acute ischemic stroke: an observational comparative study. J Neurointerv Surg. 2018; 10(12):E31–+. https://doi.org/10.1136/neurintsurg-2018-013838. PMID: 29773715

147. Gudbrandsdottir S, Frederiksen H, Hasselbalch H. Thrombopoietin-receptor agonists in haematological disorders: The Danish experience. Platelets. 2012; 23(6):423–9. https://doi.org/10.3109/09537104.2011.634931. PMID: 22185370

148. Gulmez SE, Lassen AT, Aalykke C, Dall M, Andries A, Andersen BS, et al. Do statins protect against upper gastrointestinal bleeding? Br J Clin Pharmacol. 2009; 67(4):460–5. https://doi.org/10.1111/j.1365-2125.2009.03362.x. PMID: 19371320

149. Gylvin SH, Jørgensen CC, Fink-Jensen A, Johansson PI, Kehlet H, Jørgensen CC, et al. Psychopharmacologic treatment and blood transfusion in fast-track total hip and knee arthroplasty. Transfusion.2017; 57(4):971–6. https://doi.org/10.1111/trf.13992. PMID: 28145024

150. Hallas J, Christensen R, Andersen M, Friis S, Bjerrum L. Long term use of drugs affecting the reninangiotensin system and the risk of cancer: a population-based case-control study. Br J Clin Pharmacol. 2012; 74(1):180–8. https://doi.org/10.1111/j.1365-2125.2012.04170.x. PMID: 22243442

151. Hallas J, Pottegard A. Performance of the High-dimensional Propensity Score in a Nordic Healthcare Model. Basic Clin Pharmacol Toxicol. 2017; 120(3):312–7. https://doi.org/10.1111/bcpt.12716. PMID: 27889951

152. Halmin M, Rostgaard K, Lee BK, Wikman A, Norda R, Nielsen KR, et al. Length of Storage of Red Blood Cells and Patient Survival After Blood Transfusion A Binational Cohort Study. Ann Intern Med.2017; 166(4):248–+. https://doi.org/10.7326/M16-1415. PMID: 27992899

153. Hansen CT, Kehlet H, Møller C, Mørch L, Utzon J, Ottesen B. Timing of heparin prophylaxis and bleeding complications in hysterectomy a nationwide prospective cohort study of 9,949 Danish women. Acta Obstet Gynecol Scand. 2008; 87(10):1039–47. https://doi.org/10.1080/00016340802419384. PMID: 18792841

154. Hansen MB, Jensen ML, Carstensen B. Causes of death among diabetic patients in Denmark. Diabetologia. 2012; 55(2):294–302. https://doi.org/10.1007/s00125-011-2383-2 PMID: 22127411

155. Hansen RB, Jacobsen S. Infections increase risk of arterial and venous thromboses in danish patients with systemic lupus erythematosus: 5102 Patient-years of followup. J Rheumatol. 2014; 41(9):1817–22. https://doi.org/10.3899/jrheum.131399. PMID: 25128505

156. Hansen TB, Lindholt JS, Sogaard R. Role of Experience With Preventive Medication and Personal Risk Attitude in Non-Attendance at Triple Vascular Screening. Eur J Vasc Endovasc Surg. 2018; 56(2):282–90. https://doi.org/10.1016/j.ejvs.2018.04.016. PMID: 29891436

157. Harshman LC, Xie W, Bjarnason GA, Knox JJ, MacKenzie M, Wood L, et al. Conditional survival of patients with metastatic renal-cell carcinoma treated with VEGF-targeted therapy: a population-based study. Lancet Oncol. 2012; 13(9):927–35. https://doi.org/10.1016/S1470-2045(12)70285-1. PMID: 22877847

158. Hatz C, Beck B, Steffen R, Genton B, d'Acremont V, Loutan L, et al. Real-life versus package insert: A post-marketing study on adverse-event rates of the virosomal hepatitis A vaccine Epaxal (R) in healthy travellers. Vaccine. 2011; 29(31):5000–6. https://doi.org/10.1016/j.vaccine.2011.04.099. PMID: 21569813

159. Haueis P, Russmann S, Zorina OI, Grohmann R, Kullak-Ublick GA, Jaquenoud Sirot E, et al. Coprescription of levodopa with antipsychotics in a population of 84 596 psychiatric inpatients from 1994 to 2008. Pharmacopsychiatry. 2012; 45(4):127–32. https://doi.org/10.1055/s-0031-1291348. PMID: 22086741

160. Havelin LI, Fenstad AM, Salomonsson R, Mehnert F, Furnes O, Overgaard S, et al. The Nordic Arthroplasty Register Association A unique collaboration between 3 national hip arthroplasty registries with 280,201 THRs. Acta Orthop. 2009; 80(4):393–401. https://doi.org/10.3109/17453670903039544. PMID: 19513887

161. Head SJ, Howell NJ, Osnabrugge RLJ, Bridgewater B, Keogh BE, Kinsman R, et al. The European association for cardio-thoracic surgery (EACTS)database: An introduction. Eur J Cardiothorac Surg.2013; 44(3):e175–e80. https://doi.org/10.1093/ejcts/ezt303. PMID: 23786918

162. Helgstrand F, Rosenberg J, Bay-Nielsen M, Friis-Andersen H, Wara P, Jorgensen LN, et al. Establishment and initial experiences from the Danish Ventral Hernia Database. Hernia. 2010; 14(2):131–5. https://doi.org/10.1007/s10029-009-0592-0. PMID: 19937076

163. Helgstrand F, Rosenberg J, Kehlet H, Strandfelt P, Bisgaard T. Reoperation Versus Clinical Recurrence Rate After Ventral Hernia Repair. Ann Surg. 2012; 256(6):955–8. https://doi.org/10.1097/SLA. 0b013e318254f5b9. PMID: 22580941

164. Helqvist L, Erichsen R, Gammelager H, Johansen MB, Sorensen HT. Quality of ICD-10 colorectal cancer diagnosis codes in the Danish National Registry of Patients. Eur J Cancer Care. 2012; 21(6):722– 7.

165. Helweg-Larsen K. The Danish Register of Causes of Death. Scand J Public Health. 2011; 39(7 Suppl):26–9. https://doi.org/10.1177/1403494811399958. PMID: 21775346

166. Hemkens LG, Saccilotto R, Reyes SL, Glinz D, Zumbrunn T, Grolimund O, et al. Personalized prescription feedback using routinely collected data to reduce antibiotic use in primary care a randomized clinical trial. JAMA Intern Med. 2017; 177(2):176–83. https://doi.org/10.1001/jamainternmed.2016. 8040. PMID: 28027333

167. Henningsen AK, Pinborg A, Lidegaard O, Vestergaard C, Forman JL, Andersen AN. Perinatal outcome of singleton siblings born after assisted reproductive technology and spontaneous conception: Danish national sibling-cohort study. Fertil Steril. 2011; 95(3):959–63. https://doi.org/10.1016/j. fertnstert.2010.07.1075. PMID: 20813359

168. Henningsen A-KA, Romundstad LB, Gissler M, Nygren K-G, Lidegaard O, Skjaerven R, et al. Infant and maternal health monitoring using a combined Nordic database on ART and safety. Acta Obstet Gynecol Scand. 2011; 90(7):683–91. https://doi.org/10.1111/j.1600-0412.2011.01145.x. PMID: 21477001

169. Henriksen NA, Helgstrand F, Vogt KC, Jorgensen LN, Bisgaard T, Danish Hernia D, et al. Risk factors for incisional hernia repair after aortic reconstructive surgery in a nationwide study. J Vasc Surg. 2013;57(6):1524–+. https://doi.org/10.1016/j.jvs.2012.11.119. PMID: 23548175

170. Herzberg G, Boeckstyns M, Sorensen AI, Axelsson P, Kroener K, Liverneaux P, et al. Remotion total wrist arthroplasty: Preliminary results of a prospective international multicenter

study of 215 cases. J Wrist Surg. 2012; 1(1):17–22. https://doi.org/10.1055/s-0032-1323642. PMID: 23904975

171. Hetland ML. DANBIO-powerful research database and electronic patient record. Rheumatology.2011; 50(1):69–77. https://doi.org/10.1093/rheumatology/keq309. PMID: 21148154

172. Holland-Bill L, Xu H, Sorensen HT, Acquavella J, Svaerke C, Gammelager H, et al. Positive predictive value of primary inpatient discharge, diagnoses of infection among cancer patients in the DanishNational Registry of Patients. Ann Epidemiol. 2014; 24(8):593–7. https://doi.org/10.1016/j.annepidem. 2014.05.011. PMID: 25084702

173. Horsdal HT, Mehnert F, Rungby J, Johnsen SP. Type of Preadmission Antidiabetic Treatment and Outcome among Patients with Ischemic Stroke: A Nationwide Follow-up Study. J Stroke Cerebrovasc Dis. 2012; 21(8):717–25. https://doi.org/10.1016/j.jstrokecerebrovasdis.2011.03.007. PMID: 21536457

174. Hyldig N, Joergensen JS, Wu C, Bille C, Vinter CA, Sorensen JA, et al. Cost-effectiveness of incisional negative pressure wound therapy compared with standard care after caesarean section in obese women: a trial-based economic evaluation. BJOG. 2019; 126(5):619–27. https://doi.org/10.1111/ 1471-0528.15573. PMID: 30507022

175. Ingeholm P, Gogenur I, Iversen LH. Danish Colorectal Cancer Group Database. Clin Epidemiol. 2016;8:465–8. https://doi.org/10.2147/CLEP.S99481 PMID: 27822086

176. Ittermann T, Schipf S, Dorr M, Thuesen BH, Jorgensen T, Volzke H, et al. Hyperthyroxinemia is positively associated with prevalent and incident type 2 diabetes mellitus in two population-based samples from Northeast Germany and Denmark. Nutr Metab Cardiovasc Dis. 2018; 28(2):173–9. https://doi. org/10.1016/j.numecd.2017.10.016. PMID: 29239740

177. Iversen LH, Green A, Ingeholm P, Osterlind K, Gogenur I. Improved survival of colorectal cancer in Denmark during 2001-2012-The efforts of several national initiatives. Acta Oncol. 2016; 55:10–23.https://doi.org/10.3109/0284186X.2015.1131331. PMID: 26859340

178. Jacobs L, Thijs L, Jin Y, Zannad F, Mebazaa A, Rouet P, et al. Heart 'omics' in AGEing (HOMAGE): design, research objectives and characteristics of the common database. J Biomed Res. 2014; 28(5):349–59. https://doi.org/10.7555/JBR.28.20140045. PMID: 25332706

179. Jakobsen KD, Wallach-Kildemoes H, Bruhn CH, Hashemi N, Pagsberg AK, Fink-Jensen A, et al. Adverse events in children and adolescents treated with quetiapine: an analysis of adverse drug reaction reports from the Danish Medicines Agency database. Int Clin Psychopharm. 2017; 32(2):103–6.

180. Jensen A, Sharif H, Frederiksen K, Kjaer SK. Use of fertility drugs and risk of ovarian cancer: Danish Population Based Cohort Study. BMJ. 2009; 338:b249. https://doi.org/10.1136/bmj.b249. PMID: 19196744

181. Jensen LH, Altaf R, Harling H, Jensen M, Laurberg S, Lindegaard JC, et al. Clinical outcome in 520 consecutive Danish rectal cancer patients treated with short course preoperative radiotherapy. Eur J Surg Oncol. 2010; 36(3):237–43. https://doi.org/10.1016/j.ejso.2009.10.008. PMID: 19880268

182. Jensen HI, Ammentorp J, Arding H. Withholding or withdrawing therapy in Danish regional ICUs: Frequency, patient characteristics and decision process. Acta Anaesthesiol Scand. 2011; 55(3):344–51. https://doi.org/10.1111/j.1399-6576.2010.02375.x. PMID: 21288218

183. Jensen KK, Krarup P-M, Scheike T, Jorgensen LN, Mynster T. Incisional hernias after open versus laparoscopic surgery for colonic cancer: a nationwide cohort study. Surgical Endoscopy and Other Interventional Techniques. 2016; 30(10):4469–79. https://doi.org/10.1007/s00464-016-4779-z. PMID: 26895908

184. Jensen KK, Oma E, Harling H, Krarup P-M. Type of incision does not predict abdominal wall outcome after emergency surgery for colonic anastomotic leakage. Int J Colorectal Dis. 2017; 32(6):865–73. https://doi.org/10.1007/s00384-017-2810-z. PMID: 28391448

185. Jeppesen MM, Jensen PT, Hansen DG, Iachina M, Mogensen O. The nature of early-stage endometrial cancer recurrence-A national cohort study. Eur J Cancer. 2016; 69:51–60. https://doi.org/10.1016/j.ejca.2016.09.033. PMID: 27816832

186. Johannesdottir SA, Horvath-Puho E, Ehrenstein V, Schmidt M, Pedersen L, Sorensen HT. Existing data sources for clinical epidemiology: The Danish National Database of Reimbursed Prescriptions. Clin Epidemiol. 2012; 4:303–13. https://doi.org/10.2147/CLEP.S37587. PMID: 23204870

187. Jørgensen JB, Bondeven P, Iversen LH, Laurberg S, Pedersen BG. Pelvic insufficiency fractures frequently occur following preoperative chemo-radiotherapy for rectal cancer—a nationwide MRI study. Colorectal Dis. 2018; 20(10):873–80. https://doi.org/10.1111/codi.14224. PMID: 29673038

188. Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T, et al. Directional dominance on stature and cognition in diverse human populations. Nature. 2015; 523(7561):459–62. https://doi.org/10. 1038/nature14618. PMID: 26131930

189. Kachuri L, Saarela O, Bojesen SE, Davey Smith G, Liu G, Landi MT, et al. Mendelian Randomization and mediation analysis of leukocyte telomere length and risk of lung and head and neck cancers. Int J Epidemiol. 2018.

190. Kaltoft A, Jensen LO, Maeng M, Tilsted HH, Thayssen P, Bottcher M, et al. 2-Year Clinical Outcomes After Implantation of Sirolimus-Eluting, Paclitaxel-Eluting, and Bare-Metal Coronary Stents Results From the WDHR (Western Denmark Heart Registry). J Am Coll Cardiol. 2009; 53(8):658–64. https:// doi.org/10.1016/j.jacc.2008.09.058. PMID: 19232897

191. Karkov LL, Schytte-Hansen S, Haugbolle LS. Discrepancies between sources providing the medication histories of acutely hospitalised patients. Pharm World Sci. 2010; 32(4):449–54. https://doi.org/10.1007/s11096-010-9390-0. PMID: 20446111

192. Kent P, Kongsted A, Jensen TS, Albert HB, Schiottz-Christensen B, Manniche C. SpineData—a Danish clinical registry of people with chronic back pain. Clin Epidemiol. 2015; 7:369–79. https://doi.org/10.2147/CLEP.S83830. PMID: 26316820

193. Khanna N, Opravil M, Furrer H, Cavassini M, Vernazza P, Bernasconi E, et al. CD4(+) T cell count recovery in HIV type 1-infected patients is independent of class of antiretroviral therapy. Clin Infect Dis. 2008; 47(8):1093–101. https://doi.org/10.1086/592113. PMID: 18783328

194. Khatami R, Luca G, Baumann CR, Bassetti CL, Bruni O, Canellas F, et al. The European Narcolepsy Network (EU-NN) database. J Sleep Res. 2016; 25(3):356–64. https://doi.org/10.1111/jsr.12374. PMID: 26809504

195. Kiderlen M, Bastiaannet E, Walsh PM, Keating NL, Schrodi S, Engel J, et al. Surgical treatment of early stage breast cancer in elderly: an international comparison. Breast Cancer Res Treat. 2012; 132(2):675–82. https://doi.org/10.1007/s10549-011-1892-5. PMID: 22119939

196. Kildemoes HW, Sorensen HT, Hallas J. The Danish National Prescription Registry. Scand J Public Health. 2011; 39(7 Suppl):38–41. https://doi.org/10.1177/1403494810394717. PMID: 21775349

197. Kirwan BA, Lubsen J, Brouwer Sd, van Dalen FJ, Pocock SJ, Clayton T, et al. Quality management of a large randomized double-blind multi-centre trial: The ACTION experience. Contemp Clin Trials.2008; 29(2):259–69. https://doi.org/10.1016/j.cct.2007.10.001. PMID: 18029294

198. Klein M, Gogenur I, Rosenberg J. Postoperative use of non-steroidal anti-inflammatory drugs in patients with anastomotic leakage requiring reoperation after colorectal resection: cohort study based on prospective data. BMJ. 2012;345.

199. Knudsen ST, Mosbech TH, Hansen B, Konig E, Johnsen PC, Kamper A-L. Difficulties in reaching therapeutic goals for hypertension and dysplipidaemia in patients with type 2 diabetes in general practice. Dan Med J. 2013; 60(12).

200. Kowalska JD, Friis-Moller N, Kirk O, Bannister W, Mocroft A, Sabin C, et al. The Coding Causes of Death in HIV (CoDe) Project Initial Results and Evaluation of Methodology. Epidemiology. 2011; 22 (4):516–23. https://doi.org/10.1097/EDE.0b013e31821b5332. PMID: 21522013

201. Kronborg CN, Hallas J, Jacobsen IA. Prevalence, awareness, and control of arterial hypertension in Denmark. J Am Soc Hypertens. 2009; 3(1):19–24.e2. https://doi.org/10.1016/j.jash.2008.08.001. PMID: 20409941

202. Laenkholm A-V, Jensen M-B, Eriksen JO, Buckingham W, Ferree S, Nielsen TO, et al. The ability of PAM50 risk of recurrence score to predict 10-year distant recurrence in hormone receptor-positive postmenopausal women with special histological subtypes. Acta Oncol. 2018; 57(1):44–50. https:// doi.org/10.1080/0284186X.2017.1403044. PMID: 29202609

203. Laguna MP, Beemster P, Kumar P, Klingler HC, Wyler S, Anderson C, et al. Perioperative Morbidity of Laparoscopic Cryoablation of Small Renal Masses with Ultrathin Probes: A European Multicentre Experience. Eur Urol. 2009; 56(2):355–62. https://doi.org/10.1016/j.eururo.2009.05.002. PMID: 19467771

204. Landolt K, Roessler W, Ajdacic-Gross V, Derks EM, Libiger J, Kahn RS, et al. Predictors of discontinuation of antipsychotic medication and subsequent outcomes in the European First Episode Schizophrenia Trial (EUFEST). Schizophr Res. 2016; 172(1–3):145–51. https://doi.org/10.1016/j.schres. 2016.01.046. PMID: 26922655

205. Lang PZ, Hafezi NL, Khandelwal SS, Torres-Netto EA, Hafezi F, Randleman JB. Comparative functional outcomes after corneal crosslinking using standard, accelerated, and accelerated with higher total fluence protocols. Cornea. 2019; 38(4):433–41. https://doi.org/10.1097/ICO.0000000000001878. PMID: 30681515

206. Lange J, Troelsen A, Solgaard S, Otte KS, Jensen NK, Søballe K. Cementless One-Stage Revision in Chronic Periprosthetic Hip Joint Infection. Ninety-One Percent Infection Free Survival in 56 Patients at Minimum 2-Year Follow-Up. J Arthroplasty. 2017.

207. Laouali N, Pilorget C, Cyr D, Neri M, Kaerlev L, Sabroe S, et al. Occupational exposure to organic solvents and risk of male breast cancer: a European multicenter case–control study. Scand J Work Env Hea, Supplement. 2018; 44(3):310–22.

208. Larsen JR, Siersma VD, Davidsen AS, Waldorff FB, Reventlow S, de Fine Olivarius N. The excess mortality of patients with diabetes and concurrent psychiatric illness is markedly reduced by structured personal diabetes care: A 19-year follow up of the randomized controlled study Diabetes Care in General Practice (DCGP). Gen Hosp Psychiatry. 2016; 38:42–52. https://doi.org/10.1016/j.genhosppsych. 2015.10.001. PMID: 26602087

209. Larsen TB, Skjoth F, Grove EL, Nielsen PB, Christensen TD. Effectiveness of self-managed oral anticoagulant therapy in patients with recurrent venous thromboembolism A propensity-matched cohort study. Thromb Haemostasis. 2016; 116(3):524–9.

210. Laursen PN, Holmvang L, Lonborg J, Kober L, Hofsten DE, Helqvist S, et al. Comparison between patients included in randomized controlled trials of ischemic heart disease and real-world data. A nationwide study. Am Heart J. 2018; 204:128–38. https://doi.org/10.1016/j.ahj.2018.05.018. PMID: 30103092

211. Leboeuf-Yde C, Jensen RK, Axe´n I. Absence of low back pain in patients followed weekly over one year with automated text messages. Chiropr Man Ther. 2012;20.

212. Lehnert P, Lange T, Moller CH, Olsen PS, Carlsen J. Acute Pulmonary Embolism in a National Danish Cohort: Increasing Incidence and Decreasing Mortality. Thromb Haemostasis. 2018; 118(3):539–46.

213. Lildballe DL, Vogel I, Petersen OB, Vestergaard EM. Diagnostic performance of quantitative fluorescence PCR analysis in high-risk pregnancies after combined first-trimester screening. Dan Med J. 2014; 61(11).

214. Linauskas A, Overvad K, Johansen MB, Stengaard-Pedersen K, de Thurah A. Positive predictive value of first-time rheumatoid arthritis diagnoses and their serological subtypes in the Danish National Patient Registry. Clin Epidemiol. 2018; 10:1709–20. https://doi.org/10.2147/CLEP.S175406. PMID: 30538575

215. Lindhardsen J, Ahlehoff O, Gislason GH, Madsen OR, Olesen JB, Torp-Pedersen C, et al. The risk of myocardial infarction in rheumatoid arthritis and diabetes mellitus: a Danish nationwide cohort study. Ann Rheum Dis. 2011; 70(6):929–34. https://doi.org/10.1136/ard.2010.143396. PMID: 21389043

216. Lindhardsen J, Ahlehoff O, Gislason GH, Madsen OR, Olesen JB, Svendsen JH, et al. Risk of atrial fibrillation and stroke in rheumatoid arthritis: Danish nationwide cohort study. BMJ. 2012; 344:e1257. https://doi.org/10.1136/bmj.e1257. PMID: 22403267

217. Linnet KM, Wisborg K, Secher NJ, Hove Thomsen P, Obel C, Dalsgaard S, et al. Coffee consumption during pregnancy and the risk of hyperkinetic disorder and ADHD: a prospective cohort study. Acta Paediatr. 2009; 98(1):173–9. https://doi.org/10.1111/j.1651-2227.2008.00980.x. PMID: 18764862

218. Liu C-L, Wemmelund H, Wang Y, Liao M, Lindholt JS, Johnsen SP, et al. Asthma Associates With Human Abdominal Aortic Aneurysm and Rupture. Arteriosclerosis Thromb Vasc Biol. 2016; 36 (3):570–8.

219. Lund H, Vyberg M, Eriksen HH, Grove A, Jensen AØ, Sunde L. Hydatidiform mole: Validity of the registration in the danish national patient registry, the danish cancer registry, and the danish pathology registry 1999–2009. Clin Epidemiol. 2018; 10:1223–31. https://doi.org/10.2147/CLEP.S169657. PMID: 30271218

220. Lundstrøm LH, Moller AM, Rosenstock C, Astrup G, Gatke MR, Wetterslev J, et al. Avoidance of neuromuscular blocking agents may increase the risk of difficult tracheal intubation: a cohort study of 103812 consecutive adult patients recorded in the Danish Anaesthesia Database. Br J Anaesth. 2009;103(2):283–90. https://doi.org/10.1093/bja/aep124. PMID: 19457894

221. Luta X, Hayoz S, Krause CG, Sommerhalder K, Roos E, Strazzullo P, et al. The relationship of health/ food literacy and salt awareness to daily sodium and potassium intake among a workplace population in Switzerland. Nutr Metab Cardiovasc Dis. 2018; 28(3):270–7. https://doi.org/10.1016/j.numecd. 2017.10.028. PMID: 29310971

222. Lydiksen L, Jensen-Fangel S, Blaakaer J. Is it possible to define an optimal time for chemotherapy after surgery for ovarian cancer? Gynecol Oncol. 2014; 133(3):454–9. https://doi.org/10.1016/j.ygyno. 2014.04.004. PMID: 24726613

223. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health. 2011; 39(7 Suppl):30–3. https://doi.org/10.1177/1403494811401482. PMID: 21775347

224. Maeng M, Jensen LO, Kaltoft A, Hansen HHT, Bottcher M, Lassen JF, et al. Comparison of stent thrombosis, myocardial infarction, and mortality following drug-eluting versus bare-metal stent coronary intervention in patients with diabetes mellitus. Am J Cardiol. 2008; 102(2):165–72. https://doi.org/10.1016/j.amjcard.2008.03.034. PMID: 18602515

225. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018; 50(11):1505–13. https://doi.org/10.1038/s41588-018-0241-6. PMID: 30297969

226. Majholm B, Bartholdy J, Christoffersen JK, Engbaek J, Moller AM. Poor agreement between data from the National Patient Registry and the Danish Patient Insurance Association. Dan Med J. 2012; 59(6): A4430. PMID: 22677234

227. Mareri M, Filippetti M, Ghirardini A, Vespasiano F, Di Ciaccio P, Costa AN. The EUROCET network: Support for coding, Vigilance and Surveillance. Transfu Med Hemoth. 2011; 38(6):352–6.

228. Margulis AV, Houben E, Hallas J, Overbeek JA, Pottegard A, Torp-Pedersen T, et al. Ophthalmic nepafenac use in the Netherlands and Denmark. Acta Ophthalmol. 2017; 95(5):509–17. https://doi. org/10.1111/aos.13468. PMID: 28493461

229. May MT, Ingle SM, Costagliola D, Justice AC, de Wolf F, Cavassini M, et al. Cohort profile: Antiretroviral therapy cohort collaboration (ART-CC). Int J Epidemiol. 2014; 43(3):691–702. https://doi.org/10. 1093/ije/dyt010. PMID: 23599235

230. Mejdahl MK, Andersen KG, Gartner R, Kroman N, Kehlet H. Persistent pain and sensory disturbances after treatment for breast cancer: six year nationwide follow-up study. BMJ. 2013; 346.

231. Mellernkjær L, Steding-Jessen M, Frederiksen K, Andersson M, Ejlertsen B, Jensen M-B, et al. Risk of contralateral breast cancer after tamoxifen use among Danish women. Ann Epidemiol. 2014; 24(11):843–8. https://doi.org/10.1016/j.annepidem.2014.08.003. PMID: 25277504

232. Messerli M, Blozik E, Vriends N, Hersberger KE. Impact of a community pharmacist-led medication review on medicines use in patients on polypharmacy—a prospective randomised controlled trial. BMC Health Serv Res. 2016; 16.

233. Mikkelsen S, Krueger AJ, Zwisler ST, Brochner AC. Outcome following physician supervised prehospital resuscitation: a retrospective study. BMJ Open. 2015; 5(1).

234. Minnerup J, Trinczek B, Storck M, Hohenberger M, Wilpsba¨umer S, Abdul-Rahim AH, et al. Feasibility platform for stroke studies: An online tool to improve eligibility criteria for clinical trials. Stroke. 2015; 46(1):137–42. https://doi.org/10.1161/STROKEAHA.114.007124. PMID: 25395412

235. Modvig L, Vase M, d'Amore F. Clinical and treatment-related features determining the risk of late relapse in patients with diffuse large B-cell lymphoma. Br J Haematol. 2017; 179(1):75–82. https://doi. org/10.1111/bjh.14822. PMID: 28653321

236. Mo¨hring T, Karch A, Falk CS, Laue T, D'Antiga L, Debray D, et al. Immune status in children before liver transplantation—A cross-sectional analysis within the ChilsSFree multicentre cohort study. Front Immunol. 2019; 10(JAN).

237. Møller S, Jensen MB, Ejlertsen B, Bjerre KD, Larsen M, Hansen HB, et al. The clinical database and the treatment guidelines of the Danish Breast Cancer Cooperative Group (DBCG); its 30-years experience and future promise. Acta Oncol. 2008; 47(4):506–24. https://doi.org/10.1080/02841860802059259. PMID: 18465317

238. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. Scand J Public Health. 2011; 39(7 Suppl):54–7. https://doi.org/10.1177/1403494810395825. PMID: 21775352

239. Mortensen FV, Jensen LS, Sorensen HT, Pedersen L. Cause-specific mortality associated with leukoreduced, buffy coat-depleted, or no blood transfusion after elective surgery for colorectal cancer: a posttrial 15-year follow-up study. Transfusion. 2011; 51(2):259–63. https://doi.org/10.1111/j.15372995.2010.02825.x. PMID: 20804531

240. Mortensen HR, Jensen K, Grau C. Aspiration pneumonia in patients treated with radiotherapy for head and neck cancer. Acta Oncol. 2013; 52(2):270–6. https://doi.org/10.3109/0284186X.2012.742205. PMID: 23173758

241. Mueller DR, Schmidt SJ, Roder V. One-year randomized controlled trial and follow-up of integrated neurocognitive therapy for schizophrenia outpatients. Schizophr Bull. 2015; 41(3):604–16. https://doi. org/10.1093/schbul/sbu223. PMID: 25713462

242. Mukai TO, Bro F, Olesen F, Vedsted P. To test or not: A registry-based observational study of an online decision support for prostate-specific antigen tests. Int J Med Inform. 2013; 82(10):973–9. https://doi.org/10.1016/j.ijmedinf.2013.06.010. PMID: 23850383

243. Müller R, Cieza A, Geyh S. Rasch analysis of the Hospital Anxiety and Depression Scale in spinal cord injury. Rehabil Psychol. 2012; 57(3):214–23. https://doi.org/10.1037/a0029287. PMID: 22946609

244. Munk C, Nielsen A, Liaw KL, Kjaer SK. Genital warts in men: A large population-based cross-sectional survey of Danish men. Sex Transm Infect. 2012; 88(8):640–4. https://doi.org/10.1136/sextrans-2012050512. PMID: 22941865

245. Narath SH, Mautner SI, Svehlikova E, Schultes B, Pieber TR, Sinner FM, et al. An untargeted metabolomics approach to characterize short-term and long-term metabolic changes after bariatric surgery. PloS one. 2016; 11(9).

246. Neelon SEB, Andersen CS, Morgen CS, Kamper-Jorgensen M, Oken E, Gillman MW, et al. Early child care and obesity at 12 months of age in the Danish National Birth Cohort. Int J Obes. 2015; 39(1):33–8.

247. Nickenig G, Estevez-Loureiro R, Franzen O, Tamburino C, Vanderheyden M, Lu¨scher TF, et al. Percutaneous mitral valve edge-to-edge Repair: In-hospital results and 1-year follow-up of 628 patients of the 2011–2012 pilot European Sentinel Registry. J Am Coll Cardiol. 2014; 64(9):875–84. https://doi. org/10.1016/j.jacc.2014.06.1166. PMID: 25169171

248. Nielsen J, Nielsen RE, Correll CU. Predictors of Clozapine Response in Patients With Treatment Refractory Schizophrenia Results From a Danish Register Study. J Clin Psychopharmacol. 2012; 32(5):678–83. https://doi.org/10.1097/JCP.0b013e318267b3cd. PMID: 22926603

249. Nielsen J, Jensen SOW, Friis RB, Valentin JB, Correll CU. Comparative Effectiveness of Risperidone Long-Acting Injectable vs First-Generation Antipsychotic Long-Acting Injectables in Schizophrenia: Results From a Nationwide, Retrospective Inception Cohort Study. Schizophr Bull. 2015; 41(3):627–36. https://doi.org/10.1093/schbul/sbu128. PMID: 25180312

250. Nielsen PB, Larsen TB, Skjoth F, Gorst-Rasmussen A, Rasmussen LH, Lip GYH. Restarting Anticoagulant Treatment After Intracranial Hemorrhage in Patients With Atrial Fibrillation and the Impact on Recurrent Stroke, Mortality, and Bleeding A Nationwide Cohort Study. Circulation. 2015; 132(6):517–25. https://doi.org/10.1161/CIRCULATIONAHA.115.015735. PMID: 26059010

251. Nielsen SF, Nordestgaard BG. Negative statin-related news stories decrease statin persistence and increase myocardial infarction and cardiovascular mortality: a nationwide prospective cohort study. Eur Heart J. 2016; 37(11):908–16. https://doi.org/10.1093/eurheartj/ehv641. PMID: 26643266

252. Nilsson H, Grove EL, Larsen TB, Nielsen PB, Skjoth F, Maegaard M, et al. Sex Differences in Treatment Quality of Self-Managed Oral Anticoagulant Therapy: 6,900 Patient-Years of Follow-Up. PloS one. 2014; 9(11).

253. Nolan-Kenney RC, Liu M, Akhand O, Calabresi PA, Paul F, Petzold A, et al. Optimal intereye difference thresholds by optical coherence tomography in multiple sclerosis: An international study. Ann Neurol. 2019; 85(5):618–29. https://doi.org/10.1002/ana.25462. PMID: 30851125

254. Nørskov AK, Rosenstock CV, Wetterslev J, Astrup G, Afshari A, Lundstrom LH. Diagnostic accuracy of anaesthesiologists' prediction of difficult airway management in daily clinical

practice: a cohort study of 188 064 patients registered in the Danish Anaesthesia Database. Anaesthesia. 2015; 70(3):272–81. https://doi.org/10.1111/anae.12955. PMID: 25511370

255. Nørskov AK, Wetterslev J, Rosenstock CV, Afshari A, Astrup G, Jakobsen JC, et al. Prediction of difficult mask ventilation using a systematic assessment of risk factors vs. existing practice–a cluster randomised clinical trial in 94,006 patients. Anaesthesia. 2017; 72(3):296–308. https://doi.org/10.1111/ anae.13701. PMID: 27882541

256. Nyholm AM, Gromov K, Palm H, Brix M, Kallemose T, Troelsen A. Time to Surgery Is Associated with Thirty-Day and Ninety-Day Mortality After Proximal Femoral Fracture: A Retrospective Observational Study on Prospectively Collected Data from the Danish Fracture Database Collaborators. J Bone Joint Surg Am. 2015; 97(16):1333–9. https://doi.org/10.2106/JBJS.O.00029. PMID: 26290084

257. Olsen JH, Friis S, Poulsen AH, Fryzek J, Harving H, Tjonneland A, et al. Use of NSAIDs, smoking and lung cancer risk. Br J Cancer. 2008; 98(1):232–7. https://doi.org/10.1038/sj.bjc.6604151. PMID: 18087276

258. Olsen J, Berdeaux G, Skov J. Glaucoma costs in Denmark in treatment naive patients. Acta Ophthalmol. 2013; 91(1):25–31. https://doi.org/10.1111/j.1755-3768.2011.02212.x. PMID: 21834919

259. Orsted DD, Bojesen SE, Nielsen SF, Nordestgaard BG. Association of Clinical Benign Prostate Hyperplasia with Prostate Cancer Incidence and Mortality Revisited: A Nationwide Cohort Study of 3 009 258 Men. Eur Urol. 2011; 60(4):691–8. https://doi.org/10.1016/j.eururo.2011.06.016. PMID: 21705134

260. Ozcan C, Juel K, Flensted Lassen J, von Kappelgaard LM, Mortensen PE, Gislason G. The Danish Heart Registry. Clin Epidemiol. 2016; 8:503–8. https://doi.org/10.2147/CLEP.S99475. PMID: 27822091

261. Pacurariu AC, Straus SM, Trifirò G, Schuemie MJ, Gini R, Herings R, et al. Useful Interplay Between Spontaneous ADR Reports and Electronic Healthcare Records in Signal Detection. Drug Saf. 2015;38(12):1201–10. https://doi.org/10.1007/s40264-015-0341-5. PMID: 26370104

262. Pagh A, Vedtofte T, Lynggaard CD, Rubek N, Lonka M, Johansen J, et al. The value of routine followup after treatment for head and neck cancer. A National Survey from DAHANCA. Acta Oncol. 2013;52(2):277–84. https://doi.org/10.3109/0284186X.2012.741324 PMID: 23320772

263. Palnum KH, Mehnert F, Andersen G, Ingeman A, Krog BR, Bartels PD, et al. Use of Secondary Medical Prophylaxis and Clinical Outcome Among Patients With Ischemic Stroke A

Nationwide Follow-Up Study. Stroke. 2012; 43(3):802–+. https://doi.org/10.1161/STROKEAHA.111.635342. PMID: 22207506

264. Pasternak B, Svanstrom H, Melbye M, Hviid A. Association of Treatment With Carvedilol vs Metoprolol Succinate and Mortality in Patients With Heart Failure. JAMA Intern Med. 2014; 174(10):1597–604. https://doi.org/10.1001/jamainternmed.2014.3258. PMID: 25173681

265. Patadia VK, Schuemie MJ, Coloma PM, Herings R, van der Lei J, Sturkenboom M, et al. Can electronic health records databases complement spontaneous reporting system databases? A historicalreconstruction of the association of rofecoxib and acute myocardial infarction. Front Pharmacol. 2018; 9(JUN).

266. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun. 2016; 7.

267. Paulsen MS, Andersen M, Thomsen JL, Schroll H, Larsen PV, Lykkegaard J, et al. Multimorbidity and Blood Pressure Control in 37 651 Hypertensive Patients From Danish General Practice. J Am Heart Assoc. 2013; 2(1).

268. Pechmann A, Koenig K, Bernert G, Schachtrup K, Schara U, Schorling D, et al. SMArtCARE—A platform to collect real-life outcome data of patients with spinal muscular atrophy. Orphanet J Rare Dis. 2019; 14.

269. Pedersen N, Duricova D, Lenicek M, Elkjaer M, Bortlik M, Andersen PS, et al. Infliximab dependency is related to decreased surgical rates in adult Crohn's disease patients. Eur J Gastroenterol Hepatol. 2010; 22(10):1196–203. https://doi.org/10.1097/MEG.0b013e32833dde2e. PMID: 20739896

270. Pedersen CB. The Danish Civil Registration System. Scand J Public Health. 2011; 39:22–5. https:// doi.org/10.1177/1403494810387965. PMID: 21775345

271. Pedersen EG, Hallas J, Hansen K, Jensen PEH, Gaist D. Identifying Patients with Myasthenia for Epidemiological Research by Linkage of Automated Registers. Neuroepidemiology. 2011; 37(2):120–8. https://doi.org/10.1159/000331481. PMID: 21986220

272. Perera G, Pedersen L, Ansel D, Alexander M, Arrighi HM, Avillach P, et al. Dementia prevalence and incidence in a federation of European Electronic Health Record databases: The European Medical Informatics Framework resource. Alzheimers Dement. 2018; 14(2):130–9. https://doi.org/10.1016/j. jalz.2017.06.2270. PMID: 28734783

273. Perregaard H, Aronson JK, Dalhoff K, Hellebek A. Medication errors detected in non-traditional databases: Types of errors in methotrexate dosing as listed in four different Danish registers. Eur J Clin Pharmacol. 2015; 71(11):1375–9. https://doi.org/10.1007/s00228-015-1910-3. PMID: 26257248

274. Petersen KU, Pedersen JE, Bonde JP, Ebbehoj NE, Hansen J. Mortality in a cohort of Danish firefighters; 1970–2014. Int Arch Occup Environ Health. 2018; 91(6):759–66. https://doi.org/10.1007/s00420018-1323-6. PMID: 29808435

275. Petersen PB, Kehlet H, Jorgensen CC, Lundbeck Fdn Ctr Fast Track H. Myocardial infarction following fast-track total hip and knee arthroplasty-incidence, time course, and risk factors: a prospective cohort study of 24,862 procedures. Acta Orthop. 2018; 89(6):603–9. https://doi.org/10.1080/17453674.2018. 1517487. PMID: 30326755

276. Piazza N, Wenaweser P, van Gameren M, Pilgrim T, Tsikas A, Otten A, et al. Relationship between the logistic EuroSCORE and the Society of Thoracic Surgeons Predicted Risk of Mortality score in patients implanted with the CoreValve ReValving System-A Bern-Rotterdam Study. Am Heart J. 2010;159(2).

277. Piltoft JS, Larsen SB, Dalton SO, Johansen C, Baker JL, Cederkvist L, et al. Early life risk factors for testicular cancer: a case-cohort study based on the Copenhagen School Health Records Register. Acta Oncol. 2017; 56(2):220–4. https://doi.org/10.1080/0284186X.2016.1266085. PMID: 28080187

278. Pinborg A, Ortoft G, Loft A, Rasmussen SC, Ingerslev HJ. Cervical conization doubles the risk of preterm and very preterm birth in assisted reproductive technology twin pregnancies. Hum Reprod. 2015;30(1):197–204. https://doi.org/10.1093/humrep/deu260. PMID: 25358346

279. Pironi L, Konrad D, Brandt C, Joly F, Wanten G, Agostini F, et al. Clinical classification of adult patients with chronic intestinal failure due to benign disease: An international multicenter cross-sectional survey. Clin Nutr. 2017.

280. Plüss-Suard C, Pannatier A, Kronenberg A, Mu¨hlemann K, Zanetti G. Impact of antibiotic use on carbapenem resistance in Pseudomonas aeruginosa: Is there a role for antibiotic diversity? Antimicrob Agents Chemother. 2013; 57(4):1709–13. https://doi.org/10.1128/AAC.01348-12 PMID: 23357763

281. Pommergaard HC, Klein M, Burcharth J, Rosenberg J, Dahl JB. Variation in postoperative non-steroidal anti-inflammatory analgesic use after colorectal surgery: A database analysis. BMC Anesthesiol.2014;14. https://doi.org/10.1186/1471-2253-14-14

282. Pottegard A, Christensen R, Houji A, Christiansen CB, Paulsen MS, Thomsen JL, et al. Primary nonadherence in general practice: a Danish register study. Eur J Clin Pharmacol. 2014; 70(6):757–63. https://doi.org/10.1007/s00228-014-1677-y. PMID: 24756147

283. Pottegard A, Garcia Rodriguez LA, Poulsen FR, Hallas J, Gaist D. Antithrombotic drugs and subarachnoid haemorrhage risk. A nationwide case-control study in Denmark. Thromb Haemost. 2015; 114(5):1064–75. https://doi.org/10.1160/TH15-04-0316. PMID: 26202836

284. Poulsen E, Christensen HW, Overgaard S, Hartvigsen J. Prevalence of hip osteoarthritis in chiropractic practice in Denmark: A descriptive cross-sectional and prospective study. Journal of Manipulative and Physiological Therapeutics. 2012; 35(4):263–71. https://doi.org/10.1016/j.jmpt.2012.01.010. PMID: 22417795

285. Poulsen FR, Halle B, Pottegard A, Garcia Rodriguez LA, Hallas J, Gaist D. Subdural hematoma cases identified through a Danish patient register: diagnosis validity, clinical characteristics, and preadmission antithrombotic drug use. Pharmacoepidem Dr S. 2016; 25(11):1253–62.

286. Poulsen LO, Yilmaz MK, Ljungmann K, Jespersen N, Wille-Jorgensen P, Petersen LN, et al. Local recurrence rate in a national Danish patient cohort after curative treatment for rectal cancer. Acta Oncol. 2018; 57(12):1639–45. https://doi.org/10.1080/0284186X.2018.1497299. PMID: 30169998

287. Preston MA, Riis AH, Ehrenstein V, Breau RH, Batista JL, Olumi AF, et al. Metformin Use and Prostate Cancer Risk. Eur Urol. 2014; 66(6):1012–20. https://doi.org/10.1016/j.eururo.2014.04.027. PMID: 24857538

288. Prins BP, Mead TJ, Brody JA, Sveinbjornsson G, Ntalla I, Bihlmeyer NA, et al. Exome-chip meta-analysis identifies novel loci associated with cardiac conduction, including ADAMTS6. Genome Biol. 2018; 19(1).

289. Pukkala E, Martinsen JI, Lynge E, Gunnarsdottir HK, Sparen P, Tryggvadottir L, et al. Occupation and cancer—follow-up of 15 million people in five Nordic countries. Acta Oncol. 2009; 48(5):646–790. https://doi.org/10.1080/02841860902913546. PMID: 19925375

290. Radovanovic D, Erne P. AMIS Plus: Swiss registry of acute coronary syndrome. Heart. 2010; 96(12):917–21. https://doi.org/10.1136/hrt.2009.192302. PMID: 20538666

291. Ramlau-Hansen CH, Olesen AV, Parner ET, Sorensen HT, Olsen J. Perinatal markers of estrogen exposure and risk of testicular cancer: follow-up of 1,333,873 Danish males born between 1950 and 2002. Cancer Causes Control. 2009; 20(9):1587–92. https://doi.org/10.1007/s10552-009-9403-2. PMID: 19636804

292. Rasmussen M, Ammari ABH, Pedersen B, Tønnesen H. Smoking cessation intervention activities and outcome before, during and after the national Healthcare Reform in Denmark. Clin Health Promot. 2012; 2:26–35.

293. Rasmussen M, Tønnesen H. The Danish smoking cessation database. Clin Health Promot. 2016. 6(2):36–41.

294. Rasmussen M, Fernandez E, Tonnesen H. Effectiveness of the Gold Standard Programme compared with other smoking cessation interventions in Denmark: a cohort study. BMJ Open. 2017; 7(2).

295. Rathe JO. The effect of generic switching on concerns about medicine and non-persistence among Danish adults in a general practice setting. Dan Med J. 2015; 62(10).

296. Reyes C, Pottegard A, Schwarz P, Javaid MK, Van Staa TP, Cooper C, et al. Real-Life and RCT Participants: Alendronate Users Versus FITs' Trial Eligibility Criterion. Calcif Tissue Int. 2016; 99(3):243– 9. https://doi.org/10.1007/s00223-016-0141-7. PMID: 27099132

297. Ringdal KG, Lossius HM, Jones JM, Lauritsen JM, Coats TJ, Palmer CS, et al. Collecting core data in severely injured patients using a consensus trauma template: An international multicentre study. Critical Care. 2011; 15(5).

298. Roberto G, Leal I, Sattar N, Loomis AK, Avillach P, Egger P, et al. Identifying cases of type 2 diabetes in heterogeneous data sources: Strategy from the EMIF project. PloS one. 2016; 11(8).

299. Rudin C, Burri M, Shen Y, Rode R, Nadal D, Pediat Infectious Dis Grp S, et al. Long-term safety and effectiveness of ritonavir, nelfinavir, and lopinavir/ritonavir in antiretroviral-experienced HIV-infected children. Pediatr Infect Dis J. 2008; 27(5):431–7. https://doi.org/10.1097/INF.0b013e3181646d5a. PMID: 18382386

300. Rungby J, Schou M, Warrer P, Ytte L, Andersen GS. Prevalence of cardiovascular disease and evaluation of standard of care in type 2 diabetes: a nationwide study in primary care. Cardiovasc Endocrinol.2017; 6(4):145–51. https://doi.org/10.1097/XCE.0000000000000135. PMID: 29276653

301. Russell RJ, Chachi L, FitzGerald JM, Backer V, Olivenstein R, Titlestad IL, et al. Effect of tralokinumab, an interleukin-13 neutralising monoclonal antibody, on eosinophilic airway inflammation in uncontrolled moderate-to-severe asthma (MESOS): a multicentre, double-blind, randomised, placebo-controlled phase 2 trial. Lancet Respir Med. 2018; 6(7):499–510. https://doi.org/10.1016/S22132600(18)30201-7. PMID: 29793857

302. Schaefer HH, de Villiers JD, Lotze U, Sivukhina E, Burnier M, Noll G, et al. Patients with Non-Insulin Depending Diabetes Mellitus and Metabolic Syndrome Are Suboptimal Treated in

Swiss Primary Care. Clin Exp Hypertens. 2013; 35(7):496–505. https://doi.org/10.3109/10641963.2012.758275. PMID: 23301512

303. Scha¨fer R, Strnad V, Polga´r C, Uter W, Hildebrandt G, Ott OJ, et al. Quality-of-life results for accelerated partial breast irradiation with interstitial brachytherapy versus whole-breast irradiation in early breast cancer after breast-conserving surgery (GEC-ESTRO): 5-year results of a randomised, phase 3 trial. Lancet Oncol. 2018; 19(6):834–44. https://doi.org/10.1016/S1470-2045(18)30195-5. PMID: 29695348

304. Schatlo B, Fung C, Fathi A-R, Sailer M, Winkler K, Daniel RT, et al. Introducing a nationwide registry: the Swiss study on aneurysmal subarachnoid haemorrhage (Swiss SOS). Acta Neurochir. 2012; 154(12):2173–8. https://doi.org/10.1007/s00701-012-1500-4. PMID: 23053275

305. Schatorje´ EJH, Gathmann B, Van Hout RWNM, De Vries E, Alsina L, Baumann U, et al. The PedPAD study: Boys predominate in the hypogammaglobulinaemia registry of the ESID online database. Clin Exp Immunol. 2014; 176(3):387–93. https://doi.org/10.1111/cei.12281 PMID: 24506305

306. Schmaal L, Hibar DP, Sa¨mann PG, Hall GB, Baune BT, Jahanshad N, et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. Mol Psychiatr. 2017; 22(6):900–9.

307. Schmidt M, Maeng M, Jakobsen CJ, Madsen M, Thuesen L, Nielsen PH, et al. Existing data sources for clinical epidemiology: The Western Denmark Heart Registry. Clin Epidemiol. 2010; 2:137–44. https://doi.org/10.2147/clep.s10190. PMID: 20865111

308. Schmidt M, Antonsen S, Hansen B, Moller J, Thordal C, Sorensen HT. Mortality following acute medical admission in Denmark: a feasibility study. Clin Epidemiol. 2010; 2:195–203. https://doi.org/10.2147/CLEP.S12171. PMID: 21042551

309. Schmidt M, Pedersen L, Maeng M, Lassen JF, Lash TL, Nielsen TT, et al. Nonsteroidal Antiinflammatory Drug Use and Cardiovascular Risks After Coronary Stent Implantation. Pharmacotherapy. 2011;31(5):458–68. https://doi.org/10.1592/phco.31.5.458. PMID: 21923427

310. Schmidt M, Jacobsen JB, Lash TL, Botker HE, Sorensen HT. 25 year trends in first time hospitalisation for acute myocardial infarction, subsequent short and long term mortality, and the prognostic impact of sex and comorbidity: a Danish nationwide cohort study. BMJ. 2012; 344:e356. https://doi.org/10.1136/ bmj.e356. PMID: 22279115

311. Schmidt M, Horvath-Puho E, Thomsen RW, Smeeth L, Sorensen HT. Acute infections and venous thromboembolism. J Intern Med. 2012; 271(6):608–18. https://doi.org/10.1111/j.1365-2796.2011. 02473.x. PMID: 22026462

312. Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology.Eur J Epidemiol. 2014; 29(8):541–9. https://doi.org/10.1007/s10654-014-9930-3. PMID: 24965263

313. Schmidt M, Maeng M, Madsen M, Sorensen HT, Jensen LO, Jakobsen C-J. The Western Denmark Heart Registry Its Influence on Cardiovascular Patient Care. J Am Coll Cardiol. 2018; 71(11):1259–72. https://doi.org/10.1016/j.jacc.2017.10.110. PMID: 29544611

314. Schneeberger D, Tyndall A, Kay J, Søndergaard KH, Carreira PE, Morgiel E, et al. Systemic sclerosis without antinuclear antibodies or Raynaud's phenomenon: A multicentre study in the prospective EULAR scleroderma trials and research (EUSTAR) database. Rheumatology. 2013; 52(3):560–7.https://doi.org/10.1093/rheumatology/kes315. PMID: 23221323

315. Schoos MM, Sejersten M, Baber U, Treschow PM, Madsen M, Hvelplund A, et al. Outcomes of Patients Calling Emergency Medical Services for Suspected Acute Cardiovascular Disease. Am J Cardiol. 2015; 115(1):13–20. https://doi.org/10.1016/j.amjcard.2014.09.042. PMID: 25456866

316. Schroll H, Thomsen JL, Andersen M, Friborg S, Søndergaard J. The danish model for improvement of diabetes care in general practice: impact of automated collection and feedback of patient data. International Journal of Family Medicine. 2012; 2012.

317. Schuemie MJ, Coloma PM, Straatman H, Herings RMC, Trifirò G, Matthews JN, et al. Using electronic health care records for drug safety signal detection: A comparative evaluation of statistical methods. Medical Care. 2012; 50(10):890–7. https://doi.org/10.1097/MLR.0b013e31825f63bf. PMID: 22929992

318. Sejbaek CS, Hageman I, Pinborg A, Hougaard CO, Schmidt L. Incidence of depression and influence of depression on the number of treatment cycles and births in a national cohort of 42 880 women treated with ART. Hum Reprod. 2013; 28(4):1100–9. https://doi.org/10.1093/humrep/des442. PMID: 23300199

319. Skyum F, Andersen V, Chen M, Pedersen C, Mogensen CB. Infectious gastroenteritis and the need for strict contact precaution procedures in adults presenting to the emergency department: a Danish register-based study. J Hosp Infect. 2018; 98(4):391–7. https://doi.org/10.1016/j.jhin.2017.11.001. PMID: 29128345

320. Skyum F, Pedersen C, Andersen V, Chen M, Franke A, Petersen D, et al. Risk factors for contagious gastroenteritis in adult patients with diarrhoea in the emergency department—A prospective observational multicentre study. BMC Infect Dis. 2019; 19(1).

321. Soerensen AV, Donskov F, Hermann GG, Jensen NV, Petersen A, Spliid H, et al. Improved overall survival after implementation of targeted therapy for patients with metastatic renal cell carcinoma: Results from the Danish Renal Cancer Group (DARENCA) study-2. Eur J Cancer. 2014; 50(3):553–62. https://doi.org/10.1016/j.ejca.2013.10.010. PMID: 24215846

322. Sommer C, Zuccolin D, Arnera V, Schmitz N, Adolfsson P, Colombo N, et al. Building clinical trials around patients: Evaluation and comparison of decentralized and conventional site models in patients with low back pain. Contemp Clinical Trials Commun. 2018; 11:120–6.

323. Sørensen J, Davidsen M, Gudex C, Pedersen KM, Bronnum-Hansen H. Danish EQ-5D population norms. Scand J Public Health. 2009; 37(5):467–74. https://doi.org/10.1177/1403494809105286. PMID: 19535407

324. Sørensen R, Gislason G, Torp-Pedersen C, Olesen JB, Fosbol EL, Hvidtfeldt MW, et al. Dabigatran use in Danish atrial fibrillation patients in 2011: a nationwide study. BMJ Open. 2013; 3(5).

325. Spoerri A, Zwahlen M, Egger M, Bopp M. The Swiss National Cohort: a unique database for national and international researchers. Int J Public Health. 2010; 55(4):239–42. https://doi.org/10.1007/ s00038-010-0160-5. PMID: 20556473

326. Stahl Madsen M, Bakgaard N. Treatment of varicose veins in Denmark. Dan Med J. 2014; 61(10): A4929–A. PMID: 25283621

327. Steenholdt C, Brynskov J, Thomsen OO, Munck LK, Fallingborg J, Christensen LA, et al. Individualised therapy is more cost-effective than dose intensification in patients with Crohn's disease who lose response to anti-TNF treatment: a randomised, controlled trial. Gut. 2014; 63(6):919–27. https://doi. org/10.1136/gutjnl-2013-305279. PMID: 23878167

328. Stewardson AJ, Allignol A, Beyersmann J, Graves N, Schumacher M, Meyer R, et al. The health and economic burden of bloodstream infections caused by antimicrobial-susceptible and non-susceptible Enterobacteriaceae and Staphylococcus aureus in European hospitals, 2010 and 2011: A multicentre retrospective cohort study. Euro Surveill. 2016; 21(33).

329. Strasser F, Blum D, Von Moos R, Cathomas R, Ribi K, Aebi S, et al. The effect of real-time electronic monitoring of patient-reported symptoms and clinical syndromes in outpatient workflow of medical oncologists: E-MOSAIC, a multicenter cluster-randomized phase III

study (SAKK 95/06). Ann Oncol.2016; 27(2):324–32. https://doi.org/10.1093/annonc/mdv576. PMID: 26646758

330. Streit S, da Costa BR, Bauer DC, Collet TH, Weiler S, Zimmerli L, et al. Multimorbidity and quality of preventive care in Swiss university primary care cohorts. PloS one. 2014; 9(4):e96142. https://doi.org/10.1371/journal.pone.0096142. PMID: 24760077

331. Strnad V, Ott OJ, Hildebrandt G, Kauer-Dorner D, Knauerhase H, Major T, et al. 5-year results of accelerated partial breast irradiation using sole interstitial multicatheter brachytherapy versus wholebreast irradiation with boost after breast-conserving surgery for low-risk invasive and in-situ carcinoma of the female breast: A randomised, phase 3, non-inferiority trial. Lancet. 2016; 387(10015):229–38. https://doi.org/10.1016/S0140-6736(15)00471-7. PMID: 26494415

332. Stukalin I, Connor Wells J, Fraccon A, Pasini F, Porta C, Lalani AKA, et al. Fourth-line therapy in metastatic renal cell carcinoma (mRCC): Results from the international mRCC database consortium (IMDC). Kidney Cancer. 2018; 2(1):31–6.

333. Su¨rder D, Pedrazzini G, Gaemperli O, Biaggi P, Felix C, Rufibach K, et al. Predictors for efficacy of percutaneous mitral valve repair using the MitraClip system: the results of the MitraSwiss registry. Heart. 2013; 99(14):1034–40. https://doi.org/10.1136/heartjnl-2012-303105. PMID: 23343688

334. Suttorp M, Schulze P, Glauche I, Go¨hring G, Von Neuhoff N, Metzler M, et al. Front-line imatinib treatment in children and adolescents with chronic myeloid leukemia: Results from a phase III trial. Leukemia. 2018; 32(7):1657–69. https://doi.org/10.1038/s41375-018-0179-9. PMID: 29925908

335. Svendsen LB, Jensen LS, Holm J, Kofoed SC, Pilegaard H, Preisler L, et al. Differences in the pattern of anastomotic leakage after oesophagectomy in two high-volume centres. Dan Med J. 2013;60(12).

336. Talman M-LM, Rasmussen BB, Andersen J, Christensen IJ. Estrogen receptor analyses in the Danish Breast Cancer Cooperative Group. History, methods, prognosis and clinical implications. Acta Oncol. 2008; 47(4):789–94. https://doi.org/10.1080/02841860801982741. PMID: 18465350

337. Thillemann TM, Pedersen AB, Mehnert F, Johnsen SP, Soballe K. Use of diuretics and risk of implant failure after primary total hip arthroplasty: A nationwide population-based study. Bone. 2009; 45(3):499–504. https://doi.org/10.1016/j.bone.2009.04.247 PMID: 19410667

338. Thomsen RW, Riis A, Kornum JB, Christensen S, Johnsen SP, Sorensen HT. Preadmission Use of Statins and Outcomes After Hospitalization With Pneumonia—Population-Based Cohort

Study of 29900 Patients. Arch Intern Med. 2008; 168(19):2081–7. https://doi.org/10.1001/archinte.168.19.2081. PMID: 18955636

339. Thornqvist C, Gislason GH, Kober L, Jensen PF, Torp-Pedersen C, Andersson C. Body mass index and risk of perioperative cardiovascular adverse events and mortality in 34,744 Danish patients undergoing hip or knee replacement. Acta Orthop. 2014; 85(5):456–62. https://doi.org/10.3109/17453674. 2014.934184. PMID: 24954493

340. Thøstesen LM, Nissen TN, Kjærgaard J, Pihl GT, Birk NM, Benn CS, et al. Bacillus Calmette-Gue´rin immunisation at birth and morbidity among Danish children: A prospective, randomised, clinical trial. Contemp Clin Trials. 2015; 42:213–8. https://doi.org/10.1016/j.cct.2015.04.006. PMID: 25896113

341. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sorensen HT. The predictive value of ICD10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med Res Methodol. 2011; 11:83. https://doi.org/10.1186/ 1471-2288-11-83. PMID: 21619668

342. Tollånes MC, Strandberg-Larsen K, Forthun I, Petersen TG, Moster D, Andersen A-MN, et al. Cohort profile: cerebral palsy in the Norwegian and Danish birth cohorts (MOBAND-CP). BMJ open. 2016; 6(9):e012777–e. https://doi.org/10.1136/bmjopen-2016-012777. PMID: 27591025

343. Trabert B, Ness RB, Lo-Ciganic W-H, Murphy MA, Goode EL, Poole EM, et al. Aspirin, Nonaspirin Nonsteroidal Anti-inflammatory Drug, and Acetaminophen Use and Risk of Invasive Epithelial Ovarian Cancer: A Pooled Analysis in the Ovarian Cancer Association Consortium. J Nat Cancer Inst. 2014; 106(2).

344. Tutolo M, Cornu JN, Bauer RM, Ahyai S, Bozzini G, Heesakkers J, et al. Efficacy and safety of artificial urinary sphincter (AUS): Results of a large multi-institutional cohort of patients with mid-term follow-up. Neurourol Urodynam. 2019; 38(2):710–8.

345. Tvedskov TF, Jensen MB, Balslev E, Ejlertsen B, Kroman N. Stage migration after introduction of sentinel lymph node dissection in breast cancer treatment in Denmark: a nationwide study. Eur J Cancer. 2011; 47(6):872–8. https://doi.org/10.1016/j.ejca.2010.11.022. PMID: 21194922

346. Tvedskov TF, Jensen M-B, Ejlertsen B, Christiansen P, Balslev E, Kroman N. Prognostic significance of axillary dissection in breast cancer patients with micrometastases or isolated tumor cells in sentinel nodes: a nationwide study. Breast Cancer Res Treat. 2015; 153(3):599–606. https://doi.org/10.1007/ s10549-015-3560-7. PMID: 26341752

347. Ulff-Moller CJ, Svendsen AJ, Viemose LN, Jacobsen S. Concordance of autoimmune disease in a nationwide Danish systemic lupus erythematosus twin cohort. Semin Arthritis Rheum. 2018; 47(4):538–44. https://doi.org/10.1016/j.semarthrit.2017.06.007. PMID: 28755788

348. Underbjerg L, Sikjaer T, Mosekilde L, Rejnmark L. Cardiovascular and renal complications to postsurgical hypoparathyroidism: a Danish nationwide controlled historic follow-up study. J Bone Miner Res.2013; 28(11):2277–85. https://doi.org/10.1002/jbmr.1979. PMID: 23661265

349. Underbjerg L, Sikjaer T, Mosekilde L, Rejnmark L. The Epidemiology of Nonsurgical Hypoparathyroidism in Denmark: A Nationwide Case Finding Study. J Bone Miner Res. 2015; 30(9):1738–44. https:// doi.org/10.1002/jbmr.2501. PMID: 25753591

350. Ungaro RC, Limketkai BN, Jensen CB, Allin KH, Agrawal M, Ullman T, et al. Stopping 5-aminosalicylates in patients with ulcerative colitis starting biologic therapy does not increase the risk of adverse clinical outcomes: analysis of two nationwide population-based cohorts. Gut. 2019; 68(6):977–84.https://doi.org/10.1136/gutjnl-2018-317021. PMID: 30420398

351. Usvyat LA, Haviv YS, Etter M, Kooman J, Marcelli D, Marelli C, et al. The MONitoring dialysis outcomes (MONDO) initiative. Blood Purif. 2013; 35(1–3):37–48. https://doi.org/10.1159/000345179. PMID: 23343545

352. Vach W, Saxer F, Holsgaard-Larsen A, Overgaard S, Farin-Glattacker E, Bless N, et al. Rationale and methods of an observational study to support the design of a nationwide surgical registry: the MIDAS study. Swiss Med Wkly. 2018; 148.

353. Van Hedel HJA, Severini G, Scarton A, O'Brien A, Reed T, Gaebler-Spira D, et al. Advanced Robotic Therapy Integrated Centers (ARTIC): An international collaboration facilitating the application of rehabilitation technologies. J Neuroeng Rehab. 2018; 15(1).

354. van Stralen KJ, Emma F, Jager KJ, Verrina E, Schaefer F, Laube GF, et al. Improvement in the renal prognosis in nephropathic cystinosis. Clin J Am Soc Nephrol. 2011; 6(10):2485–91. https://doi.org/10. 2215/CJN.02000311. PMID: 21868618

355. Vasan SK, Rostgaard K, Majeed A, Ullum H, Titlestad KE, Pedersen OBV, et al. ABO Blood Group and Risk of Thromboembolic and Arterial Disease: A Study of 1.5 Million Blood Donors. Circulation. 2016; 133(15):1449–57. https://doi.org/10.1161/CIRCULATIONAHA.115.017563 PMID: 26939588

356. Vester-Andersen MK, Vind I, Prosberg MV, Bengtsson BG, Blixt T, Munkholm P, et al. Hospitalisation, surgical and medical recurrence rates in inflammatory bowel disease 2003-2011-A Danish populationbased cohort study. J Crohns Colitis. 2014.

357. Vest-Hansen B, Riis AH, Sorensen HT, Christiansen CF. Acute admissions to medical departments in Denmark: diagnoses and patient characteristics. Eur J Intern Med. 2014; 25(7):639–45. https://doi. org/10.1016/j.ejim.2014.06.017. PMID: 24997487

358. Viberg B, Gundtoft PH, Schonnemann J, Pedersen L, Andersen LR, Titlestad K, et al. Introduction of national guidelines for restrictive blood transfusion threshold for hip fracture patients-a consecutive cohort study based on complete follow-up in national databases. J Orthop Surg Res. 2018;13. https:// doi.org/10.1186/s13018-018-0719-z

359. Villadsen MJ, Sorensen CH, Godballe C, Nygaard B. Need for thyroidectomy in patients treated with radioactive iodide for benign thyroid disease. Dan Med Bull. 2011; 58(12).

360. Walters S, Maringe C, Butler J, Brierley JD, Rachet B, Coleman MP. Comparability of stage data in cancer registries in six countries: Lessons from the International Cancer Benchmarking Partnership. Int J Cancer. 2013; 132(3):676–85. https://doi.org/10.1002/ijc.27651. PMID: 22623157

361. Weber R, Ruppik M, Rickenbach M, Spoerri A, Furrer H, Battegay M, et al. Decreasing mortality and changing patterns of causes of death in the Swiss HIV Cohort Study. HIV Med. 2013; 14(4):195–207. https://doi.org/10.1111/j.1468-1293.2012.01051.x. PMID: 22998068

362. Weigang E, Conzelmann LO, Kallenbach K, Dapunt O, Karck M. German Registry for Acute Aortic Dissection Type A (GERAADA)—Lessons Learned from the Registry. Thorac Cardiovasc Surg. 2010;58(3):154–8. https://doi.org/10.1055/s-0029-1240806. PMID: 20376725

363. Wiegand S, Keller KM, Lob-Corzilius T, Pott W, Reinehr T, Robl M, et al. Predicting weight loss and maintenance in overweight/obese pediatric patients. Horm Res Paediatr. 2014; 82(6):380–7. https:// doi.org/10.1159/000368963 PMID: 25531074

364. Wildgaard K, Ravn J, Nikolajsen L, Jakobsen E, Jensen TS, Kehlet H. Consequences of persistent pain after lung cancer surgery: a nationwide questionnaire study. Acta Anaesthesiol Scand. 2011; 55 (1):60–8. https://doi.org/10.1111/j.1399-6576.2010.02357.x. PMID: 21077845

365. Winterfeld U, Allignol A, Panchaud A, Rothuizen LE, Merlob P, Cuppers-Maarschalkerweerd B, et al. Pregnancy outcome following maternal exposure to statins: A multicentre prospective study. BJOG.2013; 120(4):463–71. https://doi.org/10.1111/1471-0528.12066. PMID: 23194157

366. Wurtzen H, Dalton SO, Andersen KK, Elsass P, Flyger HL, Sumbundu A, et al. Who participates in a randomized trial of mindfulness-based stress reduction (MBSR) after breast

cancer? A study of factors associated with enrollment among Danish breast cancer patients. Psychooncology. 2013; 22(5):1180–5. https://doi.org/10.1002/pon.3094 PMID: 22592966

367. Ylijoki-Sorensen S, Sajantila A, Lalu K, Boggild H, Boldsen JL, Boel LW. Coding ill-defined and unknown cause of death is 13 times more frequent in Denmark than in Finland. Forensic Sci Int. 2014;244:289–94. https://doi.org/10.1016/j.forsciint.2014.09.016 PMID: 25300069

368. Zalfani J, Frøslev T, Olsen M, Ben Ghezala I, Gammelager H, Arendt JFB, et al. Positive predictive value of the international classification of diseases, 10th edition diagnosis codes for anemia caused by bleeding in the Danish National Registry of Patients. Clin Epidemiol. 2012; 4(1):327–31.

369. Zecca C, Roth S, Findling O, Perriard G, Bachmann V, Pless ML, et al. Real-life long-term effectiveness of fingolimod in Swiss patients with relapsing-remitting multiple sclerosis. Eur J Neurol. 2018; 25 (5):762–7. https://doi.org/10.1111/ene.13594 PMID: 29431876

370. Zellweger U, Bopp M, Holzer BM, Djalali S, Kaplan V. Prevalence of chronic medical conditions in Switzerland: exploring estimates validity by comparing complementary data sources. BMC Public Health. 2014; 14(1):1157.

371. Zellweger U, Junker C, Bopp M, Egger M, Spoerri A, Zwahlen M, et al. Cause of death coding in Switzerland: evaluation based on a nationwide individual linkage of mortality and hospital in-patient records. Popul Health Metr. 2019; 17.

372. Zwisler AD, Rossau HK, Nakano A, Foghmar S, Eichhorst R, Prescott E, et al. The Danish Cardiac Rehabilitation Database. Clin Epidemiol. 2016; 8:451–6. https://doi.org/10.2147/CLEP.S99502. PMID: 27822083

373. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. Eur J Hum Genet. 2018; 26(2):149–56. https://doi.org/10.1038/ s41431-017-0045-7. PMID: 29187736

374. Siu LL, Lawler M, Haussler D, Knoppers BM, Lewin J, Vis DJ, et al. Facilitating a culture of responsible and effective sharing of cancer genome data. Nat Med. 2016; 22:464. https://doi.org/10.1038/nm. 4089 PMID: 27149219

375. Copenhagen Healthcare Cluster. Better use of health data. 2018. [Internet] Available from: https://www.cphhealthtech.com/~/media/chc/downloads/pdf/data%20redder%20liv%20cases/better-use-ofhealth-data-report-english-version.ashx?la=en

376. Statistik Denmark. Step-by-step practices for national register-based research projects at Statistics Denmark. 2018. [Internet] Available from: https://www.dst.dk/en/TilSalg/Forskningsservice

377. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol.2015; 7:449–90. https://doi.org/10.2147/CLEP.S91125 PMID: 26604824

378. Federal Office of Public Health (FOPH). Directories of approved health insurers and reinsurers. 2018.[Internet] Available from: https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung/krankenversicherung-versicherer aufsicht/verzeichnissekrankenundrueckversicherer.html.

379. Basin D. Risk Analysis on Different Usages of the Swiss AHV Number. 2017. [Internet] Available from: https://www.bj.admin.ch/bj/de/home/publiservice/publikationen/externe/2017-09-27.html

380. Carr D, Littler K. Sharing Research Data to Improve Public Health: A Funder Perspective. J Empir Res Hum Res. 2015; 10(3):314–6.

381. Mladovsky P, Mossialos E, McKee M. Improving access to research data in Europe. BMJ. 2008; 336 (7639):287–8. https://doi.org/10.1136/bmj.39409.633576.BE. PMID: 18258934

382. Pisani E, Aaby P, Breugelmans JG, Carr D, Groves T, Helinski M, et al. Beyond open data: realising the health benefits of sharing data. BMJ. 2016; 355:i5295. https://doi.org/10.1136/bmj.i5295. PMID: 27758792

383. Olejaz M, Nielsen AJ, Rudkjøbing A, Birk HO, Krasnik A, Herna´ndez-Quevedo C. Health Systems in Transition. Health. 2012; 14(2).

384. Doiron D, Raina P, L'Heureux F, Fortier I. Facilitating collaborative research: Implementing a platform supporting data harmonization and pooling. Norsk Epidemiologi. 2012; 21(2). https://doi.org/10.5324/ nje.v21i2.1500

385. Gordon WJ, Catalini C. Blockchain Technology for Healthcare: Facilitating the Transition to PatientDriven Interoperability. Comput Struct Biotec. 2018; 16:224–30.

386. Wong MC, Yee KC, Nohr C. Socio-Technical Considerations for the Use of Blockchain Technology in Healthcare. Stud in Health Tech Inform. 2018; 247:636–40.

387. Mettler M. Blockchain technology in healthcare: The revolution starts here. e-Health Networking, Applications and Services (Healthcom), 2016 IEEE 18th International Conference on; 2016: IEEE.

388. Arnason V. Coding and consent: moral challenges of the database project in Iceland. Bioethics. 2004; 18(1):27–49. https://doi.org/10.1111/j.1467-8519.2004.00377.x. PMID: 15168697

389. Arnason V. Iceland. In: ten Have HAMJ, Gordijn B, editors. Handbook of Global Bioethics. Dordrecht: Springer Netherlands; 2014. p. 1141–64.

390. Helgi HH, Gibbons SMC. Certainty is Absurd: Meeting Information Security Requirements in Laws on Population Genetic Databases. Med Law Inter. 2008; 9(2):151–68.

Chapter 4: Systemic Fairness for Sharing Health Data: Perspectives From Swiss Stakeholders

Lester Darryl Geneviève[1], Andrea Martani[1], Thomas Perneger[2], Tenzin Wangmo[1], Bernice Simone Elger[1,3]

[1]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland.
[2]Division of Clinical Epidemiology, Geneva University Hospitals and University of Geneva, Geneva, Switzerland.
[3]University Center of Legal Medicine, University of Geneva, Geneva, Switzerland.

Citation: Geneviève LD, Martani A, Perneger T, Wangmo T, Elger BS. Systemic Fairness for Sharing Health Data: Perspectives From Swiss Stakeholders. Frontiers in Public Health. 2021;9. DOI: 10.3389/fpubh.2021.669463 [CC-BY 4.0]

"Despite the value of open data, most labs make no systematic effort to share data with other scientists." - Michael Nielsen

4.1 Abstract

Introduction

Health research is gradually embracing a more collectivist approach, fueled by a new movement of open science, data sharing and collaborative partnerships. However, the existence of systemic contradictions hinders the sharing of health data and such collectivist endeavor. Therefore, this qualitative study explores these systemic barriers to a fair sharing of health data from the perspectives of Swiss stakeholders.

Methods

Purposive and snowball sampling were used to recruit 48 experts active in the Swiss healthcare domain, from the research/policy-making field and those having a high position in a health data enterprise (e.g., health register, hospital IT data infrastructure or a national health data initiative). Semi-structured interviews were then conducted, audio-recorded, verbatim transcribed with identifying information removed to guarantee the anonymity of participants. A theoretical thematic analysis was then carried out to identify themes and subthemes related to the topic of systemic fairness for sharing health data.

Results

Two themes related to the topic of systemic fairness for sharing health data were identified, namely (i) the hypercompetitive environment and (ii) the legal uncertainty blocking data sharing. The theme, hypercompetitive environment was further divided into two subthemes, (i) systemic contradictions to fair data sharing and the (ii) need of fair systemic attribution mechanisms.

Discussion

From the perspectives of Swiss stakeholders, hypercompetition in the Swiss academic system is hindering the sharing of health data for secondary research purposes, with the downside effect of influencing researchers to embrace individualism for career opportunities, thereby opposing the data sharing movement. In addition, there was a perceived sense of legal uncertainty from legislations governing the sharing of health data, which adds unreasonable burdens on individual researchers, who are often unequipped to deal with such facets of their data sharing activities.

4.2 Introduction

Health research is gradually breaking off from a long tradition of relatively closed and individualistic scientific endeavors to a new movement of data sharing, open science and collaborative partnerships (1–3). This is facilitated by the fact that the healthcare domain has become a data rich environment, with the increasing availability and complexity of data sources—gradually framed under the terms big data or big biomedical data, and the need to push the healthcare agenda toward precision medicine (4–6). Given the complexity of these data sources, conventional scientific approaches fall short in providing in-depth insights on the causality of diseases and subsequently the ability to adequately tailor healthcare interventions in this data deluge (7). Indeed, in contrast to industrial settings where big data approaches are well-implemented due to inherent characteristics of datasets used (e.g., acquired secondarily, high volume, low information density and easily accessible), healthcare datasets have a different status (6).

Healthcare data are more complex, expensive, predominantly kept in silos with high information density, and under additional legal protection due to their sensitive nature, which altogether render them less accessible to benefit from big data methods (6). Nevertheless, the healthcare data for the entire world in 2020 is alleged to be around 2,314 exabytes, representing a fifteenfold increase from 2013[1]. Therefore, present-day health research is also transitioning to the fourth research paradigm, termed "data-intensive science," where new technologies and techniques are required to make sense out of this data deluge (7–9). In this new paradigm, the expected benefits in terms of improved research and healthcare outcomes can be achieved if data are shared easily in an interoperable manner, from a multitude of stakeholders, for collaborative research, diagnosis and treatment (7, 10). This implies that the sharing of data is one of the foundations of the new paradigm. However, the data sharing status in healthcare has been relatively weak (1) in comparison to other domains [e.g., genomics or astronomy (11, 12)]. For instance, one of the current data-intensive science initiative is the Global Alliance for Genomics and Health (GA4GH) (13), which was founded in 2013 and aims to promote the responsible sharing of genomic data around the world, whilst abiding to a human rights framework[2].

The presumed benefits of sharing data for clinical care and research have been extensively discussed in the scientific literature. Some of them include: (i) an increase in scientific discoveries [e.g., testing of new hypotheses (14)] and their subsequent uptake in routine clinical practice, (ii) providing better care for patients, (iii) a reduction in research waste, and (iv) the

verification and reproducibility of research findings to ensure research integrity and transparency (1, 15). In addition, one of the important tangible goals of data sharing is to provide insights on rare diseases, where data are often limited. For instance, the Matchmaker Exchange (16) is a data sharing platform that allows the discovery of genes related to rare diseases through matching algorithms. In this regard, it adopts a federated approach in which different and autonomous databases (e.g., genotype and phenotype) are connected through a standard application programming interface (16). This platform has already helped to diagnose rare diseases (17).

However, data sharing also generates challenges for both data subjects (18) and primary data collectors. For instance, the International Committee of Medical Journal Editors (ICMJE) considered data sharing as an ethical imperative for the risks taken by participants enrolling in clinical trials, and it has taken over the years a strong stance on normalizing the sharing of clinical trials data (14, 19). The ICMJE also noted the persistence of some unsettled issues such as safeguarding researchers' interests in terms of proper attribution mechanisms for sharing data. Such issues can negatively influence data sharing requirements often present in academic publishing or in registers for clinical trials (20, 21). In this perspective, the ICMJE also recommends that, whenever possible, original data collectors should be offered the opportunity to collaborate on secondary research projects if their data are being used or at the minimum, their data collection efforts have to be acknowledged[3].

In spite of increasing efforts aimed at promoting data sharing in academic and research enterprises, resulting conflicts are sometimes portrayed simplistically as a black-and-white issue, where self-interests of primary data collectors are pitted against the idea of reusing data for the public good (22). To help solve this impasse, technical aspects to ease the sharing of data have received much attention, but there has been a lack of studies capturing our understanding of the incentives or disincentives that influence stakeholders' behavior with respect to data sharing (15, 23), especially from a systemic perspective. Such limited insights can hamper the efforts to promote data sharing by neglecting some cultural peculiarities, practices and interests of these stakeholders (10, 15, 24). Furthermore, the individual behavior of stakeholders with regard to data sharing is influenced largely by systemic factors (e.g., institutional policies or practices etc.) (25).

Although incentives for researchers to share data have not been investigated in depth (1), some factors discouraging data sharing from the researchers' own perspectives have been identified. For instance, preparing and managing datasets for secondary use prove to be a time-consuming

and costly process (10, 24, 26). Original data collectors often do not possess the required knowledge to successfully carry out these tasks (27). Moreover, the competitive environment of research does not foster a data sharing culture, since advancement in the academic career is linked primarily to the number of peer-reviewed publications in high impact factor journals, rather than on the number of research datasets made available for reuse (28). Besides, there is a lack of trust in the system and a real fear of getting scooped by external researchers gaining access to the data or that the data will be misused or misinterpreted by data recipients (10, 24, 26). The current systemic incentives—such as data sharing mandates from funders, journals or governments—cannot one-sidedly solve this complex issue (29). In this regard, Whitworth advocates for a broader collaborative approach where all involved stakeholders (e.g., researchers, ethics committees, journal editors and governments) can voice their opinions and reach consensus on instilling a data sharing culture (29). This is particularly important given that unilateral data sharing mandates do not address one central ethical principle, which is fairness for the primary data collectors (28).

Fairness is simply defined as, "the quality of treating people equally or in a way that is right or reasonable"[4]. In 2015, *fairness* was one of the guiding values put forward by a Committee of the *US Institute of Medicine*, to ensure that clinical trial data are shared in a responsible manner (30). It was highlighted that all involved stakeholders (from trial participants to researchers and sponsors) have an interest to ensure that *fairness* guides data sharing activities (30). Nevertheless, perceptions of fairness from involved stakeholders might differ based on their respective interests. On the one hand, trial participants might be more concerned in ensuring that no societal groups are unfairly discriminated in reaping the health benefits brought by these clinical trials (30). On the other hand, researchers might find it not only more important and fair to protect their interests relative to the data that they collected (e.g., if secondary publications are to be expected), but also to receive safeguards and due credit for the invested efforts, time and intellectual resources once their datasets are made available for reuse (30). Indeed, from the perspectives of primary data collectors, "any unsolicited and unsanctioned use of their data shall be seen as unfair" (31). Therefore, it is important that burdens and benefits of data sharing are fairly distributed between the original data collectors, the data recipients and ultimately, the data subjects. If the *fairness* dimension toward the primary data collectors is not addressed, these might only provide datasets that fulfill the minimum quality standards required for sharing, which can be at the detriment of high quality secondary research (28).

At the same time, it is also important to consider the ethical imperative of data sharing, which "requires that data that can be used for research purposes and research results should be made available for further research use to advance the common good of scientific knowledge" (32). Therefore, timely access to health data for secondary research purposes should not be discriminatory, delayed or restricted without due justification (32). In this aspect, *fair data sharing* should not be confused with the FAIR Data principles (33). Indeed, the latter consists of four principles (i.e., Findability, Accessibility, Interoperability and Reusability) that need to be applied to not only scholarly data but also to other tools used in the generation of such research data (e.g., algorithms and workflows) for subsequent re-use in either human-driven or machine-driven initiatives (33).

In this paper, we tackle the topic of *fairness* with respect to data sharing practices by presenting the relevant findings of a qualitative study we conducted with Swiss stakeholders. We focus in particular on matters of systemic elements that impact on fairness in data sharing. These are those elements that are inherent to the overall healthcare system, including the research domain and regulatory frameworks, that affect the fair sharing of health data, rather than those elements being mainly connected to individual motivations of the original data collectors.

4.3 Materials and methods

Ethics approval and consent to participate

The present study is part of a larger project titled "advancing SMart solutions and eliminating barriers for the Acquisition, Analysis, and Sharing of Health data in Switzerland" (SMAASH). The project falls outside the scope of the *Human Research Act* (HRA)[5]—the Swiss law on medical-related research—and thus does not require ethical approval according to Swiss regulation. This was confirmed by the cantonal ethics committee in Northwest and Central Switzerland, to which the project was nevertheless submitted (reference number: EKNZ req-2017-00810). The committee commented that the project does not pose any health risks to participants and it satisfies both the general ethical and scientific requirements. Study participants were then recruited via email for semi-structured interviews and informed about the nature and objectives of the study, the expected duration of the interview, and measures that would be taken to ensure confidentiality. Participants orally agreed for the interviews to be audio-recorded so that transcripts, with no personally identifiable information, could be created for further analysis. Upon request, some participants reviewed their transcripts for accuracy prior to the start of the analysis.

Research Team and Reflexivity

The research team consisted of two PhD candidates in biomedical ethics (LDG and AM), a senior researcher (TW) and the two principal study investigators (TP and BSE). After receiving training in qualitative research and acquiring the necessary interviewing skills, LDG and AM conducted the semi-structured interviews. LDG has a background in medicine and global health, while AM in law. They were supported during the analysis of the interview transcripts by TW and BSE. TW and BSE are both established scholars in qualitative health research. TP is an expert in the assessment of health services and quality of care. Constant supervision by TW, TP, and BSE helped limit possible bias in the interpretation of the data. Since study participants were mostly experts in their respective fields and often did not have a "neutral" view on the topic, it was important for the two interviewers to adapt their epistemological position throughout the interview (from co-expert to lay person or critic and vice-versa) (34). This served to expose and challenge assumptions made between the interviewer and the interviewee(s).

Recruitment and characteristics of participants

Purposive and snowball sampling techniques were used to recruit participants for the semi-structured interviews. The eligibility criteria for participation are researchers and policymakers working in the Swiss healthcare domain or individuals, with a relatively high position, involved in the collection, curation, sharing, linkage, and management of health datasets (e.g., registries, hospital IT infrastructure, national/regional data initiatives or hospital directors). As part of the overall aims of the SMAASH project, we carried out a systematic review (35) on projects collecting, sharing, and linking health data. Through projects analyzed in the review, we were able to recruit some of our participants. The response rate was 83%. We conducted 43 semi-structured interviews with 48 participants, since four of the interviews were one-to-two (n = 3) and one-to-three (n = 1). Our study participants had rich and diverse backgrounds (see Table 1). An interview date was scheduled with the consenting participants and further information and explanation were provided prior to the start of the interview.

Table 1 Characteristics of study participants

Type of Participants (n/%)	**Language Speaking Regions in Switzerland**		
	German-speaking**** n (%)	French-speaking n (%)	Italian-speaking n (%)
Researchers* (28/ 58.3%)	17 (60.7%)	8 (28.6%)	3 (10.7%)
Person involved in politics (10/ 20.8%)**	9 (90.0%)	1 (10.0%)	0 (0.0%)
Person with a high position in either a health register/IT infrastructure/national initiative on health data* (10/ 20.8%)**	8 (80.0%)	1 (10.0%)	1 (10.0%)

**abbreviated as "R" in results section*
***abbreviated as "P" in results section*
****abbreviated as "H" in results section*
***** some cantons (i.e. states) are bilingual but are classified in **Table 1** as monolingual based on the main language spoken (e.g. Bern as German-speaking)*

Data Collection

LDG, AM, TW and BSE developed the interview guide (Appendix 1), which was then pilot tested to ensure that each question was easily accessible and understandable to a broad audience and modifications were made accordingly based on feedbacks received. As of May 2018 to September 2019, LDG and AM independently conducted the 43 semi-structured interviews. The duration per interview ranged from 38 to 131 min. The majority of the interviews were conducted in English (*n = 37*), while the remaining few were conducted either in Italian (*n = 2*), German (*n = 3*) or French (*n = 1*), based on the preferences of the interviewees. All audio recordings were treated confidentially, verbatim transcribed, but omitting details in the transcripts that could help identify the interviewee.

Data Analysis

We conducted a thematic analysis, guided by the six-step framework devised by Braun and Clarke (36). The recordings were transcribed and analyzed using the qualitative analysis software, MAXQDA (versions 18 & 20). Three members of the research team (LDG, AM, and TW) conducted a series of preliminary meetings to discuss the transcripts of the first seven interviews. The result of these meetings led to the development of a coding tree and the

identification of themes/subthemes pertinent to the overall aim of the SMAASH project, which is to identify barriers and facilitators with respect to the processing of health data (e.g., collection, sharing, and linkage activities). The coding tree was further developed and finalized following the subsequent analysis of the remaining interview transcripts. LDG and AM coded individually the remaining transcripts and concertedly discussed with TW the identification of new themes/subthemes during a last series of meetings aimed at analyzing a sample of 15 additional transcripts. When the data corpus was coded with the finalized coding tree, another meeting was held with all the authors to discuss the macro topics within the data.

In this manuscript, only the themes/subthemes related to the macro topic of systemic fairness in data sharing are considered. LDG created a dataset where data extracts pertinent or relevant to the perception of systemic fairness guiding data sharing were gathered. This dataset was then reanalyzed from a "theoretical" or deductive perspective (36). That is, LDG re-analyzed the data extracts into themes that address the topic of systemic fairness and identified two themes and two subthemes. The quotes used in the results section were edited grammatically and non-English data extracts were translated. In a subsequent meeting, the authors discussed, refined and agreed on this final set of identified themes and subthemes for this study.

4.4 Results

Matters of systemic fairness

Two themes relevant to systemic fairness in the sharing of health data emerged in our study. These are (i) *hypercompetitive environment* and (ii) *legal uncertainty blocking data sharing* (Figure 1).

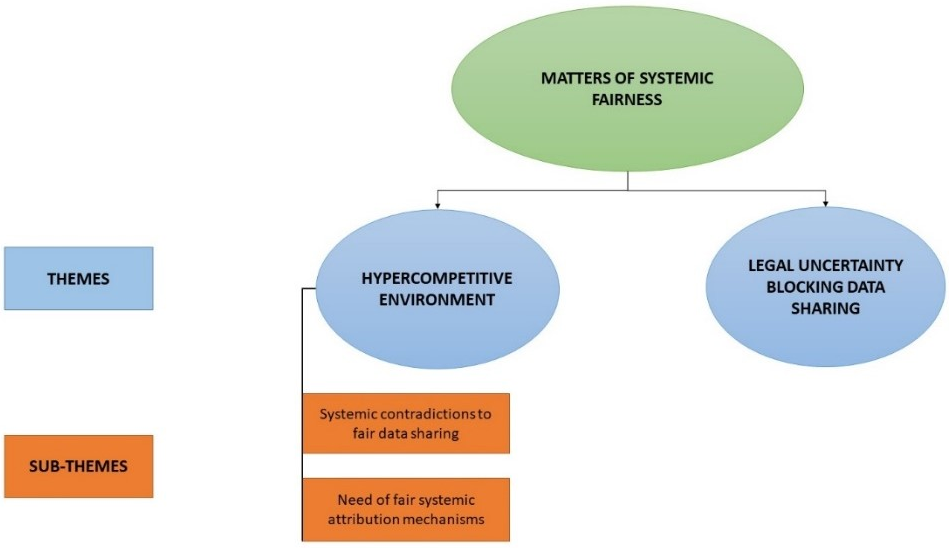


Figure 1 Themes and sub-themes on perceived systemic constraints for the fair sharing of health data

Hypercompetitive Environment

Systemic Contradictions to Fair Data Sharing

Systemic contradictions to fair data sharing comprise those barriers posed by current practices and policies, which (i) foster competition in the research context for limited resources, and (ii) hinder the fair sharing of health data. For instance, one researcher highlighted that unwillingness to share data is not simply a matter of selfishness and ego, but a failure of the system to cater for researchers' needs. She further explained that academic survival is a "real fear" that researchers have to face during their career, and it would be incorrect to attribute their protective data sharing behavior solely to their ego and self-centeredness.

> *≪I think there is a real fear, I think it is fear that they [researchers] are going to lose their spot, they are going to lose their place you know […] This is our system, I think it brings all kinds of… you know…impure intentions ultimately…you can't blame individuals. It's the system that we are working in and the system itself is a bit sick… ≫ – 43REngPA, R*

Other participants also linked the reluctance to share data with the fact that wrong incentives for career advancement are provided, which ultimately lead researchers to adopt individualistic behaviors. One participant had a relatively pessimistic view about the academic system and underlined the tremendous difficulties associated with trying to change the way it currently operates, unless there is the political will to do it.

> *≪…the credo in research is still "publish or perish" […] you need to publish papers to get additional new funds. And this system I think will not change unfortunately in the next decades. It is like…it is very difficult to change. So, if we want to change something I think you should start from the political sides or from the top one. ≫ – 2HEngPA/B, H*

The way the academic system incentivizes career advancement was perceived by some participants as being contrary to the movement of open science and data sharing. Participants stated that academic excellence is being assessed on certain quantitative performance metrics, driven primarily by publication pressure, which contrasts not only with the idea of data sharing (since researchers often end up being in competition with other researchers for publications based on their own data), but it also exposes the researcher to unfair practices (e.g., run the risk of getting *scooped*).

> *≪ The only concern we have is that the current academic system wants us to excel in our field attested by publications made in peer-reviewed journals, the issue is that we*

do not want to have projects that are then, on the pretext of sharing, which are then stolen by others» – 13RFrePA, R

« ...publish or perish is one reason, but also maybe I have some ideas, I want to test with the data first and I don't want to give others the opportunity to take away my ideas just because I didn't have the time to do it right now. So… I completely see it that you don't want to share your data the day you get them» – 22REngPA, R

Therefore, from the perspectives of some interviewees, the academic system forces researchers to take a counterintuitive inclination with regard to the timely sharing of data. For example, researchers might be unwilling to share datasets right after publication if they have not been able to fully exploit them to answer interesting research questions, in particular given the time and efforts put into collecting and curating the data. This is particularly true if there are risks that these questions could be answered quicker by external researchers, thereby depriving the original data collectors from a high quality publication:

«I mean of course we are all a bit reluctant [to share data] because we perhaps or most people perhaps share a certain fear that you know that others might use data for things that we could do ourselves and then kind of use our data to generate analysis we would like to generate at a later point in time and […]in the end it would be their publications and not ours» – 41REngPA, R

«...for many studies, you've spent four years for collecting data, and then before you would have time to analyze and publish it/and at that stage, people want the datasets because they don't spend any time on collection, they publish even the results quicker than you do. That's a reason why many researchers will be reluctant to share datasets» – 16REngPA, R

«...well first when you collect the data, you have to make a protocol to make it funded, to go through the ethics committee. So it's very big work and it needs resources, financial and human resources. So in the end, you know you've done all the work, you generated a beautiful database with everything cleaned, and you just…you know that someone asking you: "well give me your database and we will do ten papers on your work and you will not be in the authors. You will just be maybe in the acknowledgments and that's it" » – 17REngPA, R

«Because you might be working on something still especially with big studies, that sometimes takes a long time until they find the time to analyze the data that they have.

And they put a lot of effort in collecting these data and then their hands are often bound to exploit them in a timely fashion. So I think for them that it's difficult if then other people come and just, you know, just exploit≫ – 20REngPA, R

Moreover, if quantity of publications—and not only their quality—is an important element to help advance one's own career, data sharing can then be perceived as a liability. For instance, one researcher highlighted the practice of hiring researchers with more publications.

≪…The ways we incentivize people are quite wrong. They are really pushed toward individualism and you know, you have to have the best curriculum with thousands of papers, even if you never participate really in the papers […] They will hire a person publishing more than a person publishing good, good work. Even here…≫ – 17REngPA, R

Need of Fair Systemic Attribution Mechanisms

Some participants noted that there are no fair systemic attribution mechanisms for researchers and other people involved in the management, curation and preparation of datasets for secondary use. For instance, one participant underlined the difficulty in ascertaining the reasonable and acceptable way of giving credit back to the original data collectors for their work in providing the datasets.

≪Well the question is what kind of credit, right? Because if I get data from somebody and I use it, I do have to say/ I mean I give credit by saying: "I use this data from group XYZ", so in that way I already acknowledged that somebody else put the work in it and I just do the analysis. But that might not be enough credit for a researcher≫ – 20REngPA, R

Others highlighted the need for a systemic change regarding reward mechanisms to ensure that all people involved in data sharing activities receive the rightful acknowledgment for their work. One participant reported having offered co-authorship to data managers as a compensation for their contribution in ensuring the good quality of datasets for sharing purposes.

≪Our academic system is now built on impact factors in papers and there are some other ideas of valuing some other work as well. I mean preparing a dataset is an excellent example of those people [who] never get credit. And we've been working on this [name of cohort] for years. We asked people to add to the author list the [name of cohort] study group. So we get as a reference that this is [name of cohort] data. I think

we should think about a system which gives credit to people who actually collect data, manage data and [en]sure the quality…» – *23REngPA, R*

It was also noted that there is a need for funding agencies to develop and implement attribution mechanisms that consider not only the publication record of researchers but also their data sharing activities as an additional evaluation criterion for grant approvals. Indeed, even if funders provide some sort of financial compensation for data sharing activities as part of the overall project budget, acknowledging data sharing activities of researchers is perceived as a fair incentive, in particular in terms of opportunities for career advancement.

«...it [data sharing] should be acknowledged by those who have the power, a special science foundation, funding agencies, that you sort of get some benefits or some credits» – *25REngPA, R*

«I think the [Swiss funding agency] now says: "You need to have a data plan and you need/ you can put it into the budget to produce that." The problem is that for researchers that sort of effort is not moving your career very quickly forward. So this is also/ at the end of the day we then need down the road some matrix that shows: This is a researcher that is nicely sharing data and this is a researcher that doesn't nicely share data. That sort of incentives then also need to come sooner or later. It's part of your well-standing of a researcher that you have an established record of making data available if someone asks. » – *31REngTA, R*

One researcher highlighted that although certain attribution mechanisms exist for crediting researchers who made their datasets available for re-use, these are not widely implemented in the academic system. In addition, the researcher underlined that if data receivers promise to credit the data collection efforts of the original data collectors by citing the unique persistent identifiers of datasets, it could be an incentive for the original data collectors to grant access to the data.

«It [dataset] needs an ID like a paper as well. It already exists but not a lot of people use it. Each dataset needs an ID plus each dataset needs meta-data which describe the data and there should be the name of the author or the research team or the institute or whoever […] If it's possible you give access to the data to new research teams given that they reference you as the source of the data. » – *23REngPA, R*

Legal Uncertainty Blocking Data Sharing

Legal uncertainty resulting from the complex and fragmented legal landscape can be at odds with data sharing, since it can delay or even prevent access to datasets that would have otherwise been accessible for secondary use.

Some participants expressed frustration on how multi-cantonal or nationwide research projects and registries were particularly vulnerable to not only different interpretations of the same piece of legislation, but also to differing data protection requirements imposed by cantonal laws. The allegedly unclear legal situation resulted in a fragmented data sharing climate, where researchers' motivations for sharing health data were perceived as being obstructed by uncertainty resulting not only in ascertaining which data protection requirement should prevail over the others (e.g., federal or cantonal), but also by the myriad of cantonal data protection regulations. For instance, one researcher found it difficult to know under which circumstances cantonal or federal law should prevail, highlighting the complexity to navigate between these different legislations. When referring to multi-cantonal projects, the same researcher expressed difficulties in ascertaining under what conditions a specific cantonal law would apply.

> *≪Which legal bases apply in which context and what is the overlap? There is also a national law, there are cantonal laws, and so what law applies to which situation? … I'm not a lawyer, am I? ≫ – 35RGerPA, R*

> *≪So there are at least three [participating] cantons. Then there is always still the question of which cantonal law is applicable, right? Is [it] the one of the research site? Is [it] the one where we collect the data? […] So yes, we now do it the way we do it, but if a lawyer looked at it properly, she might have questions about it, right? I do not know ≫ – 35RGerPA, R*

As a consequence of the fragmented legal landscape, another researcher, with experience in managing a health database, expressed annoyance regarding the disparities in evaluation of the same research project by different cantonal data protection officers.

> *≪Yes first to the situation in Switzerland: I realized that some… It's not everything just black and white. And I realized that some [cantonal data protection officers] are very liberal and some others are not at all. I realized that there are some differences how they actually look at a specific research project or registry. So sometimes it's rather difficult to know what we are allowed to do…≫ – 23REngPA, R*

Such uncertainties had also an impact on the scope of research projects. For instance, one researcher with some experience in managing several disease registries, was frustrated when she tried to navigate through the complex fragmented legal and regulatory landscape for her different projects.

> *≪…this unclear legal situation with many questions […] one is the ethics committee do this way, and the other this way for the same thing […] the lawyers from the Federal Office of Statistics have different interpretation of the same laws as the lawyers from the Federal Office of Health, and the lawyers from the ethics committees and of the hospitals… this is complicated and usually then you can choose to do what all the lawyers agreed, which is the minimal≫ – 16REngPA, R*

Some participants also experienced difficulties determining whether their data sharing activities were legally compliant. One participant talked about the legal uncertainty arising from collaborations within European research projects, in particular since the application of the European General Data Protection Regulation (GDPR) in 2018. It was difficult for the participant to find out whether data processing activities within Switzerland abide to GDPR standards, which would have allowed the sharing of health data with a European partner. Faced with such legal uncertainty, he resorted to not sharing any personally identifying information with European partners.

> *≪…I'm looking forward that Switzerland actually gives me some guidelines how to do it with international data. I need to know what am I allowed to do with European data on servers in Switzerland and the other way around. What exactly am I allowed to send to a server in the EU with clinical data from Switzerland or patient data. The easiest way is to keep all identifying information in Switzerland. ≫ – 23REngPA, R*

Another researcher has had trouble determining whether the tools used in sharing datasets with internal or external partners were legally compliant (e.g., sharing datasets via email).

> *≪…Of course, we can always ask ethics committees if it's allowed but the tools we are using to share data, and this can be something like sending an email with a datasheet. It can be…you know there can be misuse of it. And nobody knows how secure it is≫ – 19REngSA, R*

It was also interesting to note that one person responsible for hospital data management had legal concerns regarding data security measures taken by the data recipients. For instance, she highlighted that collaborations between University Hospitals and Universities created a climate

of legal uncertainty due to the differing data security measures adopted by these two stakeholders.

> *≪...I mean the hospital has its extremely strong firewalls and as the (Swiss University Hospital) is a hospital and works closely together with University but it's not the same […] we do not really know how it's law confirmed that we can own and hand over data. ≫ – 14HEngPA, H*

4.5 Discussion

This qualitative study provides important policy-making insights on the perceived systemic determinants of fairness in the health data sharing ecosystem from the perspectives of Swiss stakeholders (mostly researchers). Our stakeholders were mostly in their mid- to late-career stages. These determinants include systemic constraints to the sharing of health data induced by hypercompetition in the Swiss research domain (i.e., systemic contradictions and the lack of fair attribution mechanisms for data sharing), and the perceived uncertainty arising from the legal and regulatory framework governing data sharing activities therein. The latter resulted in different interpretations of the legal and regulatory frameworks, and thus the need for stakeholders to sometimes err on the side of caution and give up on certain aspects of their respective projects or some collaborative research activities.

The hypercompetition between researchers for limited resources and opportunities (e.g., obtaining grants from funding agencies, academic recognition and career advancement) has been extensively discussed in the scientific literature (10, 25, 37–39). Hypercompetition is generally portrayed as having both positive and negative influences on research (40). However, our participants mostly considered the negative influences of hypercompetition on the sharing of health data, underlining increasing systemic tensions between academic survival [e.g., from "publish or perish" to "funding or famine" (37)] and the movement of open science and data sharing that is being introduced also in the Swiss context[6].

In a 2019 report to the Swiss National Science Foundation (SNSF) on Open Research Data, a survey revealed that about 75 percent of Swiss researchers actually share their data (e.g., through personal request and openly on journal, webpages and repositories). Some of the reasons given for not sharing data were (i) the need to publish before data sharing is considered, (ii) lack of time to carry out data sharing activities, (iii) fear of getting their research scooped and (iv) not receiving due credit for the shared datasets (41). The need to publish before making data available for reuse was also one of the top barriers identified by Tenopir et al. (10) among scientists globally. These reasons are in line with findings of our qualitative study and could be

explained by how the academic system primarily assesses academic performance of researchers via their publication records and their ability to secure grants from funders, rather than considering their data sharing activities. In this regard, Switzerland is not an exceptional case and mirrors the academic evaluation processes of other countries (12, 38).

To reduce the negative impact of hypercompetition on data sharing, some of our participants proposed that there is a need to provide systemic attribution mechanisms as an incentive for researchers and data managers involved in data management and sharing activities. Additionally, such a measure would allow researchers and institutions to not only receive due credit for their contribution to data sharing activities but it would also be another yardstick that could provide a better evaluation of their academic performance. Indeed, although present quantitative performance metrics (e.g., number of peer-reviewed publications) were initially used as a fair basis to distribute the scarce or finite resources (e.g., grants or academic positions) to deserving researchers, they have reportedly degraded the proper functioning of the academic reward mechanism (37).

For instance, the quantity of publications, under the so-called "publication pressure," is sometimes prioritized for distribution of finite funding resources. This was also pointed out by some of our experts, highlighting its contribution to the adoption of self-centered behaviors by researchers, which subsequently hinder the timely sharing and reuse of research data. Such phenomenon can be partly explained by a reformulation of *Goodhart's* Law by Marilyn Strathern, who stated that "when a measure becomes a target, it ceases to be a good measure" (37, 42). Therefore, even if data sharing activities are to be added as an integral quantitative performance metric in Swiss academic performance evaluations and others, it is paramount to ensure that researchers and other stakeholders do not end up producing a high number of low-quality datasets just to increase their chances in securing grants or academic positions in the hypercompetitive Swiss environment.

One way data sharing activities can be encouraged without risking to turn them in an over-individualistic metric was proposed by Pierce and colleagues (12) by highlighting that it is crucial to explicitly acknowledge the scientific value of shared datasets as an additional recognition for data generators irrespective of their publications. In this regard, they proposed to attribute and link unique persistent identifiers (UPI) to both the data generators (e.g., ORCID for scientists) and the shared datasets (e.g., digital object identifiers). Citation metadata, for the original and any subsequent publication using the shared dataset (by either the data generators or external researchers), would need to include the dataset's UPI, and the citations added to

CrossRef. These would give credit to the original data collectors every time their shared dataset is being used (12). As dataset citations are neither fully adopted nor do they capture the full complexity associated with data usage, there is also the need to develop standardized data usage metrics (e.g., the number of times a dataset has been viewed or downloaded) to better capture and measure the impact of shared datasets in moving forward research (43). If such data usage and data citation/reuse metrics are widely implemented within the research arena, they will not only offer a fairer evaluation of the academic performance of researchers (44), but they will also encourage the sharing of high quality datasets. Indeed, competition in academia is likely to be inevitable in the foreseeable future, given that resources are finite. However, it becomes a liability if it turns into a race to the bottom. By steering academic competition in the right direction through adequate and fair incentives, it could give a new impetus to data sharing activities.

With regard to the Swiss regulatory framework, uncertainty in terms of legal disagreement and legal compliance were perceived by our participants, which influenced the fair sharing of health data. One of the contributing factors is that data protection regulations are elaborated broadly to cover many situations but they lack specificity when applied to a particular context (45). Furthermore, even if there are research exemptions or research-specific rules that are implemented within certain pieces of legislations (e.g., the GDPR at the European level) or the Swiss Federal Act on Data Protection (FADP), it is still unclear for researchers how to implement these exceptions in practice (46, 47). These issues were reflected in our study where participants experienced difficulties in ascertaining whether their data sharing activities were operating within legally acceptable margins. Another contributing factor to the perceived legal uncertainty is the multitude of data protection regulations present in the Swiss legal landscape and the difficulty in identifying which specific regulation would supersede the other in a particular context. In the current Swiss legislative framework, solving this legal uncertainty requires a case-by-case approach (46), but this uncertainly slows down data sharing.

As mentioned above, fair data sharing has been proposed as an essential guidance for the exchange of health data, that data be usable to produce important research findings or promote the improvement of healthcare are promptly shared (32). If the regulatory landscape does not necessarily provide an additional layer of data protection, but rather has the main result of hampering data exchange, corrective measures should be taken. This does not necessarily require changes in legislation: it could be achieved, for example, if host institutions offered researchers the necessary training, infrastructure and legally-compliant data transfer

mechanisms to ensure that they can easily determine how to operate within legally acceptable margins. One way this could be achieved uniformly across national and international institutions is through the implementation of codes of conduct or through an adequacy model, such as the adoption of data protection certification mechanisms specifically designed to facilitate the sharing and reuse of health data for research purposes, while reducing privacy and informational harm risks for data subjects (45). In this regard, policy-makers in the healthcare and research fields could learn from experiences of successfully implemented data protection certification mechanisms in other domains.

For instance, the *Asian-Pacific Economic Cooperation*'s Cross-Border Privacy Rules System (APEC CBPR) is an established data protection certification system for trade, guaranteeing the legal compliance of companies with regard to data protection while ensuring collaboration with local governments. Indeed, "certified companies and governments are working together to ensure that when personal information moves across borders, it is protected in accordance with the standards prescribed by the system's program requirements and is enforceable across participating jurisdictions"[7]. Such an approach allows a leveling-up of data protection requirements across participating organizations, while ensuring that the local legislation within which these companies are operating is respected (48, 49). The GDPR also proposes the implementation of data protection certifications under a voluntary basis (see Art. 42)[8]. Therefore, a specific data protection certification mechanism for health research could also be useful to facilitate data sharing between institutions (45) while contributing to more fair conditions and less frustrations in the data sharing process. Moreover, this approach may also reduce the risk of imposing unfair additional financial and time constraints on already resources-limited researchers and healthcare professionals in ensuring legal compliance of their data sharing activities.

Limitations

This qualitative study has several limitations. Firstly, the low reporting of benefits of hypercompetition from our interviewees may result from the way interview questions were formulated, being more prone to discuss the barriers to health data sharing. Secondly, most of our interviewees were mid-career to late-career participants, which could have not prioritized some systemic constraints to the fair sharing of health data more pertinent to early-career researchers or healthcare professionals. Thirdly, matters of individual motivations to the fair sharing of health data have not been tackled in this article because of the richness of the data and had to be analyzed separately as part of another publication within the framework of our research project. In addition to these topic specific limitations, we do not claim our work to be

generalizable to other contexts and that social desirability bias may also have played a role in the information that we received. Moreover, some of our interview sessions had to be adapted to the needs and limited availability of our interviewees, which led to some sessions being one-to-two or one-to-three interviews.

4.6 Conclusions

The open science and data sharing movements have their raison-d'être in improving and facilitating health research, but it will be difficult to fulfill such aim unless careful consideration is given to unfair systemic inconsistencies undermining such initiatives. This qualitative study has brought into light two main systemic barriers that can undermine the fair sharing of health data from the perspectives of Swiss stakeholders. First, hypercompetition in the Swiss academic system has perverted the way finite resources are distributed to stakeholders which led them to adopt individualistic behaviors and refrain from sharing datasets. Second, a perceived legal uncertainty in the complex Swiss regulatory landscape has limited the sharing of health data by imposing unfair conditions on researchers, leaving it up to individual researchers to deal with specific interpretative and implementation aspects of the different pieces of legislation, a competence that many do not possess. As long as stakeholders believe that their legitimate interests in their datasets are not fairly safeguarded by the system, data sharing will remain difficult and the objectives of open science will be hard to achieve.

Data availability statement

The de-identified dataset used to support the conclusions of this article is available upon reasonable request to the corresponding author. To prevent the re-identification of our study participants, full transcripts cannot be shared since confidentiality was guaranteed as a prerequisite for participation.

Ethics Statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author Contributions

LG and AM were involved in the data collection part of the study. All authors were involved in the analysis and in defining the themes pertinent to systemic fairness in the sharing of health data. LG wrote the first draft of the manuscript, which was then reviewed and edited by the authors. The manuscript was then finalized and approved for submission by all authors.

Funding
Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356). The funder played no role in the design of the study and collection, analysis, and interpretation of data and in the writing of the manuscript.

Conflict of Interest
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments
The authors acknowledge the financial support provided by the Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356).

Supplementary Material
The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.669463/full#supplementary-material

Appendix 1. Interview guide for Swiss stakeholders

Abbreviations
APEC CBPR, Asia-pacific economic cooperation's cross border privacy rules system; EHR, electronic health record; FADP, federal act on data protection; GDPR, general data protection regulation; HRA, human research act; ICMJE, International Committee of Medical Journal Editors; SNSF, Swiss National Science Foundation.

Footnotes
1.EMC Digital Universe and Analysis by IDC (2014): "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things," as cited in Stanford Medicine 2017 Health Trends Report: Harnessing the Power of Data in Healthcare by Stanford Medicine (p.4). URL: https://med.stanford.edu/content/dam/sm/sm-news/documents/StanfordMedicineHealthTrendsWhitePaper2017.pdf. [Accessed date: February 17, 2021].

2.Global Alliance for Genomics and Health (2021): "Enabling responsible genomic data sharing for the benefit of human health." URL: https://www.ga4gh.org/. [Accessed date: January 21, 2021].

3.ICMJE Editors (2021): "Clinical Trials." URL: http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html. [Accessed date: January 22, 2021].

4.Cambridge English Dictionary (2020): “Fairness.” URL: https://dictionary.cambridge.org/dictionary/english/fairness. [Accessed date: June 4, 2020].

5. The Federal Council (2020): “Federal Act on Research involving Human Beings.” URL: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html. [Accessed date: June 5, 2020].

6.Swiss National Science Foundation (2020): “Open Science”. URL: http://www.snf.ch/en/researchinFocus/dossiers/open-science/Pages/default.aspx. [Accessed date: November 5, 2020].

7. Asia-Pacific Economic Cooperation (2019): “What is the Cross-Border Privacy Rules System?”. URL: https://www.apec.org/About-Us/About-APEC/Fact-Sheets/What-is-the-Cross-Border-Privacy-Rules-System. [Accessed date: November 5, 2020].

8. Intersoft Consulting (n.d.): “Art. 42 GDPR Certification”. URL: https://gdpr-info.eu/art-42-gdpr/ [Accessed date: November 5, 2020].

4.7 References

1. Rowhani-Farid A, Allen M, Barnett AG. What incentives increase data sharing in health and medical research? A systematic review. Res Integr Peer Rev. (2017) 2:4. doi: 10.1186/s41073-017-0028-9

2. Sibbald SL, Kang H, Graham ID. Collaborative health research partnerships: a survey of researcher and knowledge-user attitudes and perceptions. Health Res Pol Syst. (2019) 17:92. doi: 10.1186/s12961-019-0485-3

3. Friesike S, Schildhauer T. Open science: many good resolutions, very few incentives, yet. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M, editors. Incentives and Performance: Governance of Research Organizations. Cham: Springer International Publishing (2015). p. 277–89.

4. Toga AW, Foster I, Kesselman C, Madduri R, Chard K, Deutsch EW, et al. Big biomedical data as the key resource for discovery science. J Am Med Inform Assoc. (2015) 22:1126–31. doi: 10.1093/jamia/ocv077

5. Favaretto M, De Clercq E, Schneble CO, Elger BS. What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PLoS ONE. (2020) 15:e0228987. doi: 10.1371/journal.pone.0228987

6. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From big data to precision medicine. Front Med. (2019) 6:34. doi: 10.3389/fmed.2019.00034

7. Buchan I, Winn J, Bishop C. A unified modeling approach to data-intensive healthcare. In: Hey T, Tansley S, Tolle K, editors. The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research (2009).

8. Bell G, Hey T, Szalay A. Beyond the data deluge. Science. (2009) 323:1297–8. doi: 10.1126/science.1170411

9. Hey T, Tansley S, Tolle K. The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research (2009).

10. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PLoS ONE. (2015) 10:e0134826. doi: 10.1371/journal.pone.0134826

11. Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. Data sharing in genomics — re-shaping scientific practice. Nat Rev Genet. (2009) 10:331–5. doi: 10.1038/nrg2573

12. Pierce HH, Dev A, Statham E, Bierer BE. Credit data generators for data reuse. Nature. (2019) 570:30–2. doi: 10.1038/d41586-019-01715-4

13. Knoppers BM, Thorogood AM. Ethics and big data in health. Curr Opin Syst Biol. (2017) 4:53–7. doi: 10.1016/j.coisb.2017.07.001

14. Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, et al. Sharing clinical trial data — a proposal from the international committee of medical journal editors. N Engl J Med. (2016) 374:384–6. doi: 10.1056/NEJMe1515172

15. Piwowar HA, Becich MJ, Bilofsky H, Crowley RS, on behalf of the caBIG Data Sharing and Intellectual Capital Workspace. Towards a data sharing culture: recommendations for leadership from academic health centers. PLoS Med. (2008) 5:e183. doi: 10.1371/journal.pmed.0050183

16. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The matchmaker exchange: a platform for rare disease gene discovery. Hum Mutat. (2015) 36:915–21. doi: 10.1002/humu.22858

17. Oud MM, Tuijnenburg P, Hempel M, van Vlies N, Ren Z, Ferdinandusse S, et al. Mutations in EXTL3 cause neuro-immuno-skeletal dysplasia syndrome. Am J Hum Genet. (2017) 100:281–96. doi: 10.1016/j.ajhg.2017.01.013

18. Bauchner H, Golub RM, Fontanarosa PB. Data sharing: an ethical and scientific imperative. JAMA. (2016) 315:1237–9. doi: 10.1001/jama.2016.2420

19. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials — a requirement of the international committee of medical journal editors. N Engl J Med. (2017) 376:2277–9. doi: 10.1056/NEJMe1705439

20. Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. PLoS ONE. (2009) 4:e7078. doi: 10.1371/journal.pone.0007078

21. Goldacre B, DeVito NJ, Heneghan C, Irving F, Bacon S, Fleminger J, et al. Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource. BMJ. (2018) 362:k3218. doi: 10.1136/bmj.k3218

22. Rosenbaum L. Bridging the data-sharing divide — seeing the devil in the details, not the other camp. N Engl J Med. (2017) 376:2201–3. doi: 10.1056/NEJMp1704482

23. Let data speak to data. Nature. (2005) 438:531. doi: 10.1038/438531a

24. Fecher B, Friesike S, Hebing M. What drives academic data sharing? PLoS ONE. (2015) 10:e0118053. doi: 10.1371/journal.pone.0118053

25. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PLoS ONE. (2011) 6:e21101. doi: 10.1371/journal.pone.0021101

26. Institute of Medicine Board on Health Sciences Board on Health Sciences Policy Roundtable on Translating Genomic-Based Research for Health National Cancer Policy Forum Forum on Neuroscience and Nervous System Disorders. Sharing Clinical Research Data: Workshop Summary. In: Downey AS, Olson S, editors. Washington, DC: National Academies Press (2013). p 156.

27. Popkin G. Data sharing and how it can benefit your scientific career. Nature. (2019) 569:445–7. doi: 10.1038/d41586-019-01506-x

28. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. N Engl J Med. (2017) 376:1684–7. doi: 10.1056/NEJMsb1616595

29. Whitworth J. Data sharing: reaching consensus. Bull World Health Organ. (2010) 88:468. doi: 10.2471/BLT.10.079202

30. Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC: The National Academies Press (2015). p 304.

31. Kim D. Enabling access to clinical trial data: when is unfair use fair. Chi-Kent J Intell Prop. (2014) 14:521.

32. Meier-Abt PJ, Lawrence AK, Selter L, Vayena E, Schwede T. The Swiss approach to precision medicine. Swiss Med Wkly. (2018). doi: 10.3929/ethz-b-000274911

33. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. (2016) 3:160018. doi: 10.1038/sdata.2016.18

34. Van Audenhove L. Expert Interviews and Interview Techniques for Policy Analysis. [Presentation]. Brussels: Vrije Universiteit Brussel (2007).

35. Geneviève LD, Martani A, Mallet MC, Wangmo T, Elger BS. Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: a systematic review. PLoS ONE. (2019) 14:e0226015. doi: 10.1371/journal.pone.0226015

36. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. (2006) 3:77–101. doi: 10.1191/1478088706qp063oa

37. Edwards M, Roy S. Academic research in the 21st century: maintaining scientific integrity in a climate of perverse incentives and hypercompetition. Environ Eng Sci. (2017) 34:51–61. doi: 10.1089/ees.2016.0223

38. Waaijer CJF, Teelken C, Wouters PF, van der Weijden ICM. Competition in science: links between publication pressure, grant pressure and the academic job market. High Educ Policy. (2018) 31:225–43. doi: 10.1057/s41307-017-0051-y

39. Tijdink JK, Schipper K, Bouter LM, Maclaine Pont P, de Jonge J, Smulders YM. How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers. BMJ Open. (2016) 6:e008681. doi: 10.1136/bmjopen-2015-008681

40. Anderson MS, Ronning EA, De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. Sci Eng Ethics. (2007) 13:437–61. doi: 10.1007/s11948-007-9042-5

41. von der Heyde M. Open Research Data: Landscape and Cost Analysis of Data Repositories Currently Used by the Swiss Research Community, and Requirements for the Future. Weimar: Swiss National Science Foundation (2019).

42. Strathern M. 'Improving ratings': audit in the British University system. Eur. Rev. (1997) 5:305–21.

43. Fenner M, Lowenberg D, Jones M, Needham P, Vieglais D, Abrams S, et al. Code of practice for research data usage metrics release 1. PeerJ Preprints. (2018) 6:e26505v1. doi: 10.7287/peerj.preprints.26505

44. Cousijn H, Feeney P, Lowenberg D, Presani E, Simons N. Bringing citations and usage metrics together to make data count. Data Sci. J. (2019) 18:9. doi: 10.5334/dsj-2019-009

45. Phillips M. International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR). Hum Genet. (2018) 137:575–82. doi: 10.1007/s00439-018-1919-7

46. Martani A, Egli P, Widmer M, Elger B. Data protection and biomedical research in Switzerland: setting the record straight. Swiss Med Wkly. (2020) 150: w20332. doi: 10.4414/smw.2020.20332

47. Eiss R. Confusion over Europe's data-protection law is stalling scientific progress. Nature. (2020) 584:498. doi: 10.1038/d41586-020-02454-7

48. Asia-Pacific Economic Cooperation. APEC Privacy Framework. Singapore (2015).

49. Sullivan C. EU GDPR or APEC CBPR? A comparative analysis of the approach of the EU and APEC to cross border data transfers and protection of personal data in the IoT era. Comput Law Secur Rev. (2019) 35:380–97. doi: 10.1016/j.clsr.2019.05.004

Chapter 5: Individual notions of fair data sharing from the perspectives of Swiss stakeholders

Lester Darryl Geneviève[1], Andrea Martani[1], Bernice Simone Elger[1,2], Tenzin Wangmo[1]

[1] Institute for Biomedical Ethics, University of Basel, Basel, Switzerland
[2] University Center of Legal Medicine, University of Geneva, Geneva, Switzerland

Citation: Geneviève LD, Martani A, Elger BS, Wangmo T. Individual notions of fair data sharing from the perspectives of Swiss stakeholders. BMC Health Services Research. 2021;21(1):1007. DOI: 10.1186/s12913-021-06906-2 [CC-BY 4.0]

"Remember the early days of the Net, when everything was going to be open and free, and we were all going to share information in a techno-utopia? That was great until people realized that their user data could be turned into gold. Now there are billions at stake, and nobody is playing nice anymore." – Daniel Lyons

5.1 Abstract

Background

The meaningful sharing of health data between different stakeholders is central to the advancement of science and to improve care offered to individual patients. However, it is important that the interests of individual stakeholders involved in this data sharing ecosystem are taken into account to ensure fair data sharing practices. In this regard, this qualitative study investigates such practices from the perspectives of a subset of relevant Swiss expert stakeholders, using a distributive justice lens.

Methods

Using purposive and snowball sampling methodologies, 48 expert stakeholders from the Swiss healthcare and research domains were recruited for semi-structured interviews. After the experts had consented, the interviews were audio-recorded and transcribed verbatim, but omitting identifying information to ensure confidentiality and anonymity. A thematic analysis using a deductive approach was conducted to identify fair data sharing practices for secondary research purposes. Themes and subthemes were then identified and developed during the analysis.

Results

Three distributive justice themes were identified in the data sharing negotiation processes, and these are: (i) effort, which was subcategorized into two subthemes (i.e. a claim to data reciprocity and other reciprocal advantages, and a claim to transparency on data re-use), (ii) compensation, which was subcategorized into two subthemes (i.e. a claim to an academic compensation and a claim to a financial compensation), and lastly, (iii) contribution, i.e. the significance of data contributions should be matched with a corresponding reward.

Conclusions

This qualitative study provides insights, which could inform policy-making on claims and incentives that encourage Swiss expert stakeholders to share their datasets. Importantly, several claims have been identified and justified under the basis of distributive justice principles, whilst some are more debatable and likely insufficient in justifying data sharing activities. Nonetheless, these claims should be taken seriously and discussed more broadly. Indeed, promoting health research while ensuring that healthcare systems guarantee better services, it is paramount to ensure that solutions developed are sustainable, provide fair criteria for academic careers and promote the sharing of high quality data to advance science.

5.2 Background

Over the past few decades, data sharing has become an increasingly discussed topic in the scientific literature. It has further regained impetus following the approval of the European General Data Protection Regulation (GDPR) and its enforcement in 2018 [1]. This growing interest in data sharing also stems from the fact that its potential benefits from both a clinical and research perspective are increasingly underscored. Indeed, it is claimed that data sharing: (i) promotes new scientific discoveries through the re-analysis of shared datasets to test new hypotheses; (ii) helps in building trust and transparency in scientific findings given that their reproducibility can be independently verified; (iii) limits waste in research by preventing duplication of efforts; (iv) accelerates the uptake of research findings into routine clinical practice; and (v) improves the quality of patient care [2,3,4,5]. Therefore, proponents of data sharing argue that the latter is justified under two sets of arguments, namely those of an *ethical* and *moral* nature (e.g., reducing safety risks for research participants), and those of a *scientific* and *practical* nature (e.g., ensuring the reproducibility of research findings) [5].

Data sharing can be understood as "any form of release of research [and healthcare] data for use by others" (adapted definition from [6]). In the research context, this can be achieved by either making data available on data repositories (including on the website of researchers and their institutions, or as supplementary material in scientific publications) or by accepting to fulfill requests for data of external researchers [6]. In the United States, the *open data movement* – i.e. making datasets openly available – was encouraged by the 2009 memorandum on "Transparency and Open Government" [7, 8]. It was exemplified by the subsequent launch of *Data. Gov* [7], a US online governmental platform, which is now hosting more than 300,000 freely-available datasets [9]. Recognizing the value of research datasets to advance science beyond their initial contribution, the *open data movement* has also been taken up by the scholarly data publishing ecosystem. This led to data sharing requirements as a pre-publication condition for researchers (e.g., enforced by funders and journals [10]) and the creation of online data repositories for many scientific disciplines. For instance, the *Harvard Dataverse Repository* allows researchers to "open [their] data to the general public, or restrict access and define customizable terms of use" [11].

One important aspect behind the idea of making datasets more open, is to make them re-usable outside the context in which they are initially collected (as demonstrated by the attribution of unique *Digital Object Identifiers* for future citations [11]). In this context, good data management and stewardship are deemed essential components for an effective and appropriate

re-use of scholarly datasets. In this regard, the FAIR Data Principles (Findability, Accessibility, Interoperability and Reusability) have been formulated as a guideline to ensure that digital deliverables of scientific research can "become 'first class citizens' in the scientific publication ecosystem, where the quality of the publication—and more importantly, the impact of the publication—is a function of its ability to be accurately and appropriately found, reused, and cited over time, by all stakeholders, both human and mechanical" [12].

In spite of the different ways how data can be shared, ultimately the exchange of data between different stakeholders also depends on the willingness of several stakeholders to provide access to their datasets. It is therefore important to understand the disincentives and incentives with respect to data sharing from the perspectives of individual researchers both from a systemic and individual level [13]. In fact, whereas technological solutions (e.g. availability of data repositories) might help, they are unlikely to be the *silver bullet* that would positively influence a researcher' behavior or attitude with regard to the sharing of data, especially if social or cultural aspects are neglected [6].

There are numerous reasons that undermine or hinder the sharing of health-related data for research purposes. These include among others: (i) ethical and legal barriers, such as the need to safeguard the privacy of data subjects due to the sensitive nature of health-related data; (ii) trust issues, limited expertise and time to carry out successfully data sharing activities, apprehensions regarding potential misinterpretation or misuse of shared datasets; (iii) technical barriers (e.g. differing data standards or limited data linkage capacities), (iv) the lack of systemic attribution mechanisms for shared datasets that would give credit back to the original data collectors, and lastly (v) financial barriers (i.e. preparing a dataset for sharing is a costly procedure) [13,14,15,16,17,18]. Moreover, barriers may differ based on the individual. Indeed, an early-career researcher is less likely to report data sharing issues linked to limited time availability to deposit a dataset than a middle or late-career researcher [19]. Additionally, openly sharing datasets on online repositories or as supplementary open-access files in scientific publications could expose not only the original data collectors to new threats such as data theft [20], but also research participants to additional privacy risks. Indeed, such open-access datasets – although *anonymized* - might be linked with other publicly available datasets and increase the risk of re-identifying study participants [21].

In cases where data cannot be shared on online repositories, it is possible to negotiate access with the original data collectors. For these cases, it is important to understand which data sharing practices are considered as being fair or unfair, and therefore likely to be adopted or

rejected by researchers. In this manuscript, we propose using a distributive justice lens to identify such data sharing practices. These "fair" data sharing practices should not be confused with the FAIR Data Principles described earlier, as further explained in the data analysis part of the methods section.

Hence, this qualitative study explores the individual notions of fair data sharing from the perspectives of a subgroup of relevant Swiss stakeholders through the lenses of distributive justice. We rely on interview data collected through semi-structured interviews conducted with 48 expert stakeholders, in particular those involved predominantly in health services research but also with the inclusion of other experts active in the Swiss healthcare and policy-making fields. We analyzed their views on individual notions of fairness for sharing health data. Through an analytical framework based on the concept of distributive justice, we reflected on those data sharing practices that might be perceived as just or fair and which form part of the negotiation process during which requests to share data by external researchers are considered by data collectors. Hence, this study was conducted to better understand how this negotiation process is taking place between Swiss expert stakeholders in order to tailor recommendations for the Swiss context. In other manuscripts belonging to the same project, the distinct topics of systemic fairness for the sharing of health data and data ownership were treated [13, 22].

5.3 Methods

Ethics statement

The data for this paper was collected as part of our larger research project titled "advancing SMart solutions and eliminating barriers for the Acquisition, Analysis, and Sharing of Health data in Switzerland" (SMAASH), belonging to the Swiss National Research Programme 74 on Smarter Health Care. The SMAASH project aims to identify barriers and facilitators to the processing of health data in the Swiss context, and it does not fall under the remit of the Swiss Human Research Act (HRA). Therefore, according to Swiss legislation, the SMAASH project is exempt from requiring an ethical approval. This was confirmed by the cantonal ethics committee in Northwest and Central Switzerland (file number: EKNZ req-2017-00810), which stated that general and scientific requirements were both satisfied and that the study did not entail any risk for participants' health.

Identification and recruitment of study participants

Study participants had to fulfill the following eligibility criteria for this study: (i) at the time of recruitment, they had to be working in the Swiss healthcare domain (including research), (ii) they were either policymakers or researchers or individuals responsible for the management of

health datasets, e.g. disease registries, data linkage institutions, hospital IT infrastructures or any other data initiative (national or regional). They were recruited by purposive – which also included identification of eligible participants via a systematic review [14] that was conducted in an earlier phase of the SMAASH project - and snowball sampling, whereby interviewed participants were asked to recommend additional interviewees for this study. The identified stakeholders were approached via email, and subsequently enrolled for semi-structured interviews after having been informed about, amongst others, the aims of the research project and the data security measures in place to protect their privacy and confidentiality. The recruited 48 expert stakeholders (the majority being in the middle to late stages of their careers) were categorized into three main groups: researchers (n = 28), policy-makers (n = 10) and lastly, those having a senior position in managing health databases, Information Technology infrastructures or data initiatives (n = 10). Study participants provided their verbal consent for the interviews to be audio-recorded after having been provided with consent information and the opportunity to have their additional questions answered by the interviewer. The participants were guaranteed that the resulting recordings would be transcribed verbatim for further analysis, but excluding details that could potentially reveal their identity. Accuracy of the transcripts were also checked by some interviewees who requested it.

Interview guide and data collection

As this study is part of a larger research project, the interview guide was developed by LDG, AM, BSE and TW to answer the broad research objectives of the SMAASH project. Questions in the interview guide were further enriched and refined via additional information gathered during the simultaneous conduct of the systematic review mentioned in the previous section [14]. The interview guide was then pilot-tested and modified accordingly to ensure that interview questions were clear and easily understandable (annexed). Additional probing questions were also developed for questions judged to require additional or deeper investigation. For instance, concerning the fourth question of the annexed interview guide, probing questions were developed around the interviewee's general knowledge on the current status of data sharing activities for research in the Swiss context, the challenges encountered in sharing health data and solutions that could be implemented to address them, including general recommendations on how to improve data sharing at the local and international level. Furthermore, other follow-up probes were also asked when needed during the conduct of the interviews.

The data collection period started in May 2018 and ended in September 2019. After receiving training in qualitative interviews, 43 semi-structured interviews (with duration ranging from 38

to 131 min) were conducted by LDG and AM, of which 37 interview sessions were conducted in English, whilst the remaining few (n = 6) were conducted either in French, German or Italian depending on the preferences of the study participants. The majority of the interview sessions were one-on-one interviews, with only four being either one-on-two or one-on-three, for a total of 48 expert stakeholders interviewed for this study. The audio recordings were placed in an access-protected folder on a secured server provided by the University of Basel, which was only accessible by members of the research team.

Data analysis

A thematic analysis of the transcripts was carried out using the qualitative analysis software MAXQDA (versions 18 and 20), and was inspired by Braun and Clarke's analytical approach [23]. LDG, AM and TW were involved in the preliminary analysis of a sample of the first seven transcripts that led to the development of an initial coding tree and the identification of themes and sub-themes related to the aims of the SMAASH project. In this regard, the SMAASH project aims to identify factors that influence positively or negatively the collection, sharing and linkage of health data. The initial coding tree and the list of themes were then finalized during the ongoing analysis of the remaining transcripts by the authors. The few non-English transcripts were coded and analyzed, and relevant data extracts for this study were translated by one of the authors proficient in the interview language (mother tongue or at least C1 language level).

After the data corpus was entirely coded, the authors met to discuss the main topics stemming from it. Some of the main topics included (i) data ownership issues [22], (ii) Swiss stakeholders' recommendations to improve the health data infrastructure in the Swiss context [24], (iii) systemic issues hindering the fair sharing of health data [13], and (iv) individual notions of fairness for sharing health data. For the purpose of this study, data extracts pertaining to the individual notions of fairness for sharing health data were gathered by LDG, who subsequently carried out a deductive thematic analysis [23] and identified themes and subthemes that englobed the individual notions of fair sharing of health data through a distributive justice lens.

Distributive justice can be viewed as "the fair distribution of the burdens and benefits of social cooperation among diverse persons with competing needs and claims" [25]. To better identify and evaluate fair data sharing activities between the relevant stakeholders, *desert-based principles* – a set of distributive justice principles – were used as an analytical framework. In this regard, desert-based principles can be viewed as either falling under one of these three main classifications: (i) *compensation* (e.g. the original data collectors should receive due rewards

for the financial expenses made in collecting, managing and sharing datasets), (ii) *effort* (e.g., the original data collectors should receive due rewards for the efforts they have put in collecting, managing and sharing datasets), and (iii) *contribution* (e.g., the original data collectors should receive due rewards for the significance of their contributions in collecting, managing and sharing of datasets) [26]. Therefore, any data extract that did not include notions of fair or unfair data sharing practices based on the desert-based principles were excluded.

Research team and reflexivity

At the time of the study, LDG and AM were two doctoral candidates in biomedical ethics, who received training in conducting qualitative interviews in preparation for the field work. AM has a legal background whilst LDG's background is in medicine and global health. TW, a senior researcher, and BSE, the principal investigator of the SMAASH project, have both extensive qualitative research experience, in particular concerning health research. TW and BSE continuously supervised the analysis part of the study to help limiting misinterpretation and bias due to potential preconceptions held by LDG and AM.

5.4 Results

Individual notions of fair data sharing

Three themes pertinent to the desert-based justice principles were identified in the data extracts, namely (i) *effort*, which was further categorized into two sub-themes, (ii) *compensation*, which was also subcategorized into two sub-themes, and (iii) *contribution* (Fig. 1).

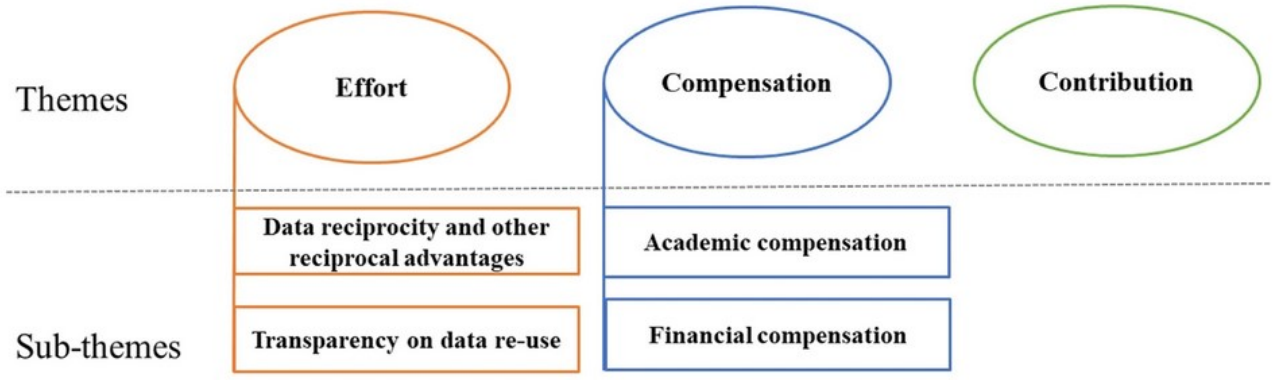


Fig. 1 Themes and sub-themes of individual notions of fair data sharing

Effort

Data reciprocity and other reciprocal advantages

We defined data reciprocity as the actions taken by the original data collector to make data available to the data recipient, only if the former receives in return additional data from the data recipient. Some participants made it clear that their main motivation for sharing data with others resides in the expectation that they shall also receive additional data from the recipients.

> "On the personal point of view, I am willing to collaborate with others, and if I share [data], I will also receive" – Res2

Others adopted a more nuanced perception of data reciprocity, highlighting the existence of mutual benefits between the original data collectors and data recipients, but also the need to tailor data sharing requirements to the efforts made by the original data collectors. For instance, one researcher responsible for the management of registries highlighted that given the efforts and resources invested by some of the original data collectors, they did not have to pay to use data from the main database.

> "And if people ask for access for data from cohorts, which I lead, […] the (medical specialty) oncologists, they have invested their time and money into collecting data, they get [data] about/of course for free" – Res8

Similarly, another stakeholder stressed the importance to cater for the needs of the original data collectors by providing them in return some valuable information that they could use. The importance of trustful relationships was noted as an important element to ensure the continued sharing of data between the original data collectors and the data recipients.

> "We share the data. We make the data useful, also for the data providers, because they can use their statistical data and show something. This contributes to build this trust and I think that it will last a few years and then most of the physicians will provide their data." – Pol4

Another person involved in health policy-making also noted that, in spite of competing interests (e.g. competition between University Hospitals), University Hospitals have agreed to share data between one another for a common objective: the advancement of the Swiss healthcare system under the framework of a national data initiative.

> "I must say they [University hospitals] realize that it is a win-win situation for everyone. And again, I don't know how much they do really share … if it's 90%, if it's only 10%, if it's 5%, I have no idea. We will see that in the future but at least, I would say those have realized that, they can't do it alone if they want still do research that's on the high level." – Pol2

Transparency on data re-use

Some study participants discussed transparency with relation to the secondary use and management of the shared datasets in return for their contribution, which was also deemed as an essential component for trustful collaborations. Furthermore, transparency was also

considered important to ensure that appropriate data security measures are taken to maintain confidentiality of the shared datasets, and to ascertain clear data ownership and data usage rules.

> "But to be in a position to collaborate we need trust. And this is// the first thing we are trying to do is to show transparently what we are doing, what is happening with the data. And to build some sort of a trustful collaboration with the providers of data." – Pol4

> "… share something if you are aware of what the other partner would do [with] the data or intends to do and how they ensure confidentiality" – Res2

> "That's why I think it remains important to really be open and clear and transparent which data belongs to whom and how can it be used." – Pol8

Because of the duty to protect the patients included in the dataset, one researcher requested that data recipients need to provide detailed and clear information on the duration that the shared datasets will remain in their possession, and that the shared datasets should not be passed further on to any third party.

> "[I would share data] With conditions, absolutely. Sure first of all, my first priority is to those patients, I did not inform them in the informed consent. I did not say: "oh I'm going to be passing your data to X, Y, Z you know". I didn't feel the need to you know…So it would have to be very clear who am I giving these data to? What are their purposes? How long these data will be available to that person? And you know whether that person can then pass them on further? That I wouldn't like, I wouldn't be comfortable with that." – Res26

Although a lot of researchers requested transparency of future data use in return for their data contribution, this was in practice operationalized by setting up contractual agreements between the original data collectors and the data recipients. For instance, some stakeholders described that the first step in the data sharing process was to initiate a contract with all clauses pertaining to the secondary use of data.

> "… the big step was negotiating a contract. Talking about details, what we are allowed to do, who has access, how are the data stored, where are they stored etc. and things like that. Once we had the contract it was rather simple. They just gave us the data, we knew what we are allowed to do, that was simple" – Res13

"… we need to sign a contract between the one who is sharing the data and the one who will use it. First to make sure that all the steps of the research project are respected." – Res9

"If I share data with you, we need to have a contract about what's happen with this data" – Stak4

Indeed, contractual agreements were considered essential for the sharing of health data because they set clear rules on how the shared datasets should be used, and they also provide some protection to the original data collectors. Contractual agreements helped ensure that (i) the shared datasets are less likely to be misused or used illegally (particularly important for potentially identifiable information), (ii) appropriate data security measures had been taken by the data recipients, (iii) publications rights are respected (e.g. avoiding competition for publications or authorship order between original data collectors and recipients), and lastly, (iv) that there is a general acceptation of the rules and conditions of use by both parties involved.

"Well in terms of researchers I would say yes, they have fear that others could publish the data but we address this with the data transfer use agreement. So that has clauses for publication rights, IT-rights and if both parties agree to that then this is solved. On a hospital level actually there is a so-called "Rahmenvertrag" [German for 'framework agreement'] where all the hospitals agreed to do what they can in order to make data available and share it with each other. So I see the intention is there, that they want to reach this, to come to a data-sharing situation" – Stak7

"That's should be some legal contract between the data sharer and the team which will use it. And of course, there should be some agreement between the two parts on publications, citations… well that's already important." – Res9

"I must also say I think even if it looks silly but I think it's really good before you share data that you have a contract where it is clear who owns the data, where does it go, who is first author, who is last author … " – Pol2

Importantly, it was further highlighted that contractual data sharing agreements put into broad daylight the responsibilities that each party has with regard to the shared dataset. Therefore, any breach of the contractual agreement's rules would be considered irresponsible action.

> "With a contract…. there is a reciprocal acceptation of the rules: what happens with the data; what shouldn't or couldn't happen with the data. And then it's a matter of responsibility." – Pol4

> "Again here I would welcome that we have sort of templates what we need to sign to share data. So I still want something in written. You know it's not just via email I send you a link to a dropbox and here you have the data, do whatever you like! […] There is a sort of a responsibility on both sides, those who make data available and those who then take on board data from others. They also have to sign something I think." – Res17

Compensation

Academic compensation

The sub-theme *academic compensation* refers to the justification of data sharing activities under the conditions that co-authorships in resulting publications or collaborations will be offered in return for the resources and work invested by the original data collectors in collecting and managing datasets. For instance, some researchers described a case where they requested data from a health insurer. In return for the data, the latter requested to be a co-author in subsequent publications, and had the upper hand over the researchers in the negotiation process. One of the researchers found the practice of requesting co-authorship as relatively surprising from a person who is not even competing academically for limited resources (e.g. career and funding opportunities) but could potentially be explained by the insurer's prior role as an academic researcher. Another interviewee justified the fairness of the health insurer's actions by referring to the resources the latter invested in the data sharing activities.

> "Participant number 3: … there was a power game around the actual data. For example, he said, "I want to be a co-author of the publications." To say, which is surprising because one imagines that he is an administrative [person working for health insurance companies], no? That is, an official. Instead well, if he is an administrative but high-level - he probably also had a doctorate - he would also have had publications […]
>
> Participant number 1: … No, well we do it, but it is a job that takes work [preparing datasets for sharing], it takes time, so we have to have something to return. [speaking of the health insurance point of view]" – Res16a/b/c

Similarly, another researcher demanded to be offered co-authorship in publications that made use of his shared dataset. Interestingly, the researcher also highlighted that the practice of offering co-authorship for shared datasets is more common for certain disciplines. Consequently, the interviewee expressed some frustrations with regard to what is current

practice in the discipline of economics, i.e. an acknowledgment for sharing data, which the interviewee deemed as insufficient.

> "Participant: The only way you could do that is that you say: "Ok we share the data but we are on each paper that is published with this data" and I think in the medicine that's more common. But in economics if you just collected the data, you end up in an acknowledgement footnote. And you are never (emphasized) a co-author.
>
> Interviewer: So your condition for sharing the data would be that you are put as a co-author?
>
> Participant: Yes. That would be one way." – Res12

Financial compensation

This type of compensation refers to the claim that data recipients should cover at least part of the expenses incurred by the original data collectors in the collection, management and/or sharing of datasets. Indeed, financial compensation was often mentioned by our interviewees as a way of receiving recognition for the collection or processing of datasets.

> "… this [data sharing] would have to be done also with some kind of financial compensation" – Res10

Moreover, one researcher highlighted that the monetary value of datasets is gradually being recognized, especially from the perspectives of healthcare professionals. For example, healthcare professionals are now asking for a financial compensation for datasets that they have collected.

> "In fact, when we arrived and said, "Give it to us," it didn't seem true. Aside from that they also sold it at a high price. But oh well we had the money and we gave it to him. So - how to say - there is a desire - I say on the part of doctors, therefore health professionals - to produce data and to use them because there is a value, a personal gain." – Res16a/b/c

Similarly, another expert stakeholder also advocated for a more in-depth discussion to clarify whether there is a need to monetize datasets produced by hospital staff. This was deemed particularly important since the latter often do not receive any benefit for the work done in collecting and managing datasets, which are subsequently used to answer research questions.

> "… people in the hospital are collecting data don't have benefits from this work, other than okay you are a good boy, good data collection from you. I think we are missing

> something there […] the secondary use of data under/monetization of the secondary use of data is something that need to be clarified." – Stak8

Going beyond fairness considerations, some researchers also highlighted that financial compensation is necessary not only for the resources invested by the original data collectors in collecting, managing and sharing datasets, but also as a means of ensuring the sustainability of some long-term projects (e.g. in the case of cohort data), and the quality of datasets. One researcher also reported that such practice is not well implemented in Switzerland but seems to be common in other countries.

> "And then if that group has money, I think they should pay something towards [data sharing] […] if it goes towards a group which can get the grant application and share a little of that for data acquisition. I mean we have done that, I've asked for a dataset from a cohort study in England to compare our own results, so we had two cohorts and they will be asked for ten or twenty thousand pounds for instance to own the data. So that is a common use I think in some other countries" – Res8.

Similarly, another researcher made an analogy to data sharing requirements imposed by Swiss Federal institutions (e.g. Federal Statistical Office), whereby researchers need to pay to get access to data. The interviewee highlighted that this is a fair practice that could be implemented by researchers who own datasets.

> "You see, when you see that some private companies are selling their data, I already bought them, and even the public one, for example the Federal Office of Statistics is selling some data and in fact, they are selling the work they invested in extraction. So it's quite fair in fact. How will they pay the people doing the work? And it's the same for us. If you want to have very good databases and with data quality management, you have to put some resources on that. So if you want to share the data, it should be funded within the project or funded by the others asking for your data." – Res9

Contribution

Contribution here refers to the importance of the original data collectors' data sharing activities for the attainment of the objectives set by the secondary research project, and that such contribution should be compensated by a matching reward made to the original data collectors. For instance, one researcher explained that original data collectors were made co-authors on a resulting manuscript and authorship order was determined based on the volume of patient data each co-author has contributed.

> "Indeed, I also had to put them as authors even if they did not even write an article .... but not even an article. But since they gave me the patients, they became authors. And the order of the authors depended on the number of patients. That is, you gave me a lot, you are first." – Res16a/b/c

Regarding data sharing between different partners for a specific research project, one researcher highlighted that for data sharing to take place, it is necessary that each partner makes more or less an equivalent data contribution. Otherwise, it could be perceived as prejudicial to those collaborators who have contributed more data, in particular if their larger invested resources and efforts are masked behind the collective endeavor.

> « … you have to realize that if you want to share the data, everyone needs to have the data and everyone needs to have more or less the same data. Otherwise, as it is very competitive, you will not necessarily want to share from a personal and scientific point of view what you have made a particular effort for, you have obtained the funds, you have found people, you have made a research project and you don't necessarily want what was a huge job for you to become ... part of a bigger project on the same topic where you lose all personal effort » - Res6

5.5 Discussion

This qualitative study explores the individual notions of fair data sharing from the perspectives of Swiss expert stakeholders, through the lens of distributive justice as an analytical framework. In this regard, this study provides insights on fair or unfair data sharing practices that form part of the negotiation process occurring within the Swiss context, whereby data collectors are requested to provide their data to external researchers. From a distributive justice point of view, individual notions of fairness were justified under the efforts made by the individual researchers in collecting, managing and sharing datasets, and these include (i) a claim to data reciprocity and other reciprocal advantages, and (ii) a claim to transparency regarding data-reuse (which was often operationalized through data transfer contractual agreements). Secondly, individual notions of fairness were also justified under the *compensation* claim, and these include: (i) a claim to an academic compensation in the form of co-authorship on manuscript publications or collaboration opportunities, and (ii) a claim to a financial compensation, that is, the costs of data sharing activities should be at minimum covered by those requesting the data. Lastly, under the *contribution* claim, the significance of the efforts (in terms of data) generated by the original data collectors were mentioned, in particular within multi-partner research collaborations (e.g. inequitable data sharing between the parties involved).

Reciprocity has usually been put forward in national and international ethical frameworks governing data sharing activities [27,28,29]. In our study, the participants hinted that receiving some reciprocal advantage in return for their contribution was a way to acknowledge their efforts in collecting and managing datasets. Such a view might be related to the competitive Swiss academic environment, where original data collectors might fear not receiving due credit for the efforts put in collecting, managing and sharing their datasets (in terms of unsatisfactory attribution mechanisms or diminishing career opportunities) [13]. Therefore, obtaining additional or equivalent data from recipients represents a motivation to share data without and at the same time losing a competitive advantage as compared to the data receivers.

Another individual motivation put forward by the original data collectors for sharing data was that of having transparency rules on the re-use and storage modalities of the shared datasets by data recipients. Transparency has been extensively discussed in many data sharing frameworks, often in the form of data availability statements for external researchers to confirm or refute the validity and test the reproducibility of certain research findings [30,31,32]. In contrast, our study shows that there is another dimension to transparency. Indeed, data collectors feel entitled to know all modalities associated with the re-use of their datasets, in the same way as data recipients are entitled to know all modalities associated with the creation of these datasets. Indeed, many interviewees stated that they would request, in exchange for sharing their datasets, that data recipients provide clear information on the intended secondary use, duration of storage, data security measures taken and a reciprocal acceptation of data ownership rules by both parties. These transparency claims are operationalized in practice by legally-binding contractual data transfer/data access agreements, where additional requirements were also stipulated (e.g. publication rights). These agreements are common practice whenever data sharing occurs through a controlled access [30].

Concerning academic compensation, receiving co-authorship opportunities as part of data sharing activities has long been perceived as a fair practice in many scientific fields, especially among researchers. For instance, in an international survey on data sharing perceptions and practices by scientists worldwide, it was observed that almost 60% of scientists (out of 1257) found that it was fair for them to receive co-authorship in exchange of their data, and 61% (out of 1226) found it fair to give co-authorship to the original data collectors if they are using their data. Furthermore, offering collaboration opportunities to the original data collectors was also perceived as fair by the great majority of scientists (81%) in return for using their data [18], a practice also recommended by the *International Committee of Medical Journal Editors*

(ICMJE) [33]. These perceptions seem to be also reflected in our qualitative study, whereby co-authorship and collaboration opportunities were mentioned as individual motivations to the sharing of data to compensate for the efforts made by data collectors. However, the question arises whether the simple act of sharing datasets solely justifies co-authorship, as it was suggested by some of our interviewees.

In that regard, the ICMJE also provides a series of four authorship criteria that need to be fulfilled for original data collectors to be given co-authorship in a scientific publication, and refutes the claim that efforts made in the acquisition of data are sufficient on their own to justify co-authorship [34]. One could think of contributorship as another means of receiving credit for the data contribution, in particular if all criteria for authorship are not satisfied. In this regard, Richard Horton once argued that contributorship might help to dismantle inappropriate authorship practices (e.g. scientists receiving undue credit), but it cannot help the scientific community to "find an appropriate, consistent, and reproducible means of judging academic merit" [35]. One promising proposition was made by Bierer, Crosas and Pierce in the form of *data authorship* [36].

It is important to differentiate between "data author" and "data collector", a term that has been extensively used in this manuscript. The difference between the two resides in the fact that any member of a research team who contributes to the data collection process is considered as an original data collector, but to *qualify* as a data author, one needs to make "substantial contributions to the original acquisition, quality control, and curation of the data, be accountable for all aspects of the accuracy and integrity of the data provided, and ensure that the available data set follows FAIR Guiding Principles" [36]. Moreover, data authorship needs also to be differentiated from manuscript authorship in the sense that data authors are only responsible for the scientific integrity of datasets, but they cannot be held accountable for the content or conclusions of a resulting manuscript, unless they are also listed as its authors. For data authorship to matter, it needs to be well-received and implemented within the scientific community, and recognized by academic institutions, journals, funding and governmental agencies as an additional criterion to reward deserving scientists for their data sharing efforts (e.g. just like the Hirsch-index is currently used in academic evaluations, a "d-index" could be envisaged for data authors) [36]. Moreover, from a distributive justice perspective, data authors could also be offered not only collaboration opportunities, but also opportunities to contribute in a substantial way to future publications in order to obtain co-authorship. This would

constitute a fair practice given the efforts and resources they have invested in making the datasets shareable whilst assuming full responsibility for their integrity.

Financial compensation was another discussed aspect, also falling under the perceived individual motivations of data sharing. In our study, financial compensation was perceived as a proper reward to cover part of the costs incurred by the original data collectors in the processing of their respective datasets but also as a means of ensuring the quality and sustainability of the health databases. Our study findings are partly aligned with the scientific literature. For example, in an international survey investigating the perceptions and practices of scientists with regard to data sharing, almost 70% of their participants rejected the idea that it is a fair for data recipients to pay the original data collectors for their datasets [18]. In contrast, Cole and colleagues [37] argued that, in the context of academic medical centers, it is fair to provide financial compensation to cover the costs of data acquisition, but stressed that it is crucial that data sharing is not being promoted for financial gains. Indeed, if financial compensation for data sharing will be widely implemented in academia, it is important that the negotiation process between the original data collectors and data recipients to be fair and transparent, with the primary objective of recovering the costs of data processing incurred by data collectors. Therefore, there should be some clear and consistent criteria on how to calculate the minimum fee that data recipients would have to pay to access data, so that the latter are not disadvantaged during the negotiation process with the downside effect of hindering data sharing. In this regard, one could learn from the data access cost calculation methods employed by NHS (National Health Service) Digital, who provide data not only to researchers but also to clinicians and commissioners for the improvement of NHS services [38].

Under the contribution principle, mechanisms to compensate for the quantity and importance of data contributions made by the original data collectors were rarely discussed. Some of our interviewees explained that in some cases, the original data collectors explicitly stated that their degree of contribution to data sharing activities needed to be matched with a corresponding reward (e.g. in the form of authorship order in publications derived from the shared datasets). Others stated that they will be reluctant to engage in data sharing activities if all research collaborators did not contribute an equivalent amount of data. In contrast to certain academic disciplines where authorship order has no value (e.g. in economics or mathematics where authors are listed in alphabetical order), in health research and other disciplines authorship order plays a central role in defining the specific contributions made by each author and also for the

academic reward mechanism - in particular how promotion in academia is often linked to the number of first- or last-author publications [39].

Therefore, in health research, authorship order aligns with distributive justice principles, since each author's contribution "should be assigned a proportionate and fair share of the overall value of the publication" [39]. Nonetheless, defining precisely the contribution of each author to the value of a manuscript is a challenging task and consequently, authorship order based on the amount of data contributed by each research collaborator offers a weak assessment of their academic merit [39]. Additionally, irrespective of their data contributions, all authors need to take full responsibility for the entire content of the manuscript whilst guaranteeing its scientific integrity (see ICMJE criteria for authorship [34]), which further undermines the claim that authorship order is dependent on the amount of data contributed by each author.

Limitations

This qualitative study has some limitations. Firstly, the findings of this study might be biased by the seniority of our study participants. Indeed, the majority of our participants were either late- or middle-career stakeholders, and therefore individual notions related to fair data sharing from the perspectives of early-career participants might have been underrepresented or neglected. Secondly, our interview sessions had to be adapted to the needs and expectations of our expert stakeholders, some of whom preferred to have one-on-two or one-on-three interviews rather than the common one-on-one interview. By their nature, group interviews are more subject to group dynamics, which could have influenced the reporting of elements deemed more important for the group of interviewees rather than for the individual. Thirdly, the development of our interview guide was informed by issues related to data sharing that are more predominant in the Swiss context (e.g. those identified in our systematic review [14]). Fourthly, we also acknowledge that this study offers only an initial assessment of the issues identified in the Swiss context (some being blurry concepts) and needed further investigation. Therefore, these issues were explored, clarified and discussed with relevant expert stakeholders during a follow-up Delphi-based workshop [40]. Finally, like every other qualitative study, our findings are not generalizable and may be affected by social desirability bias, where participants may have tended to discuss expected ethical concerns.

5.6 Conclusions

This qualitative study provides insights that could inform policy-making on individual motivations that need to be accounted for to promote fair data sharing from the perspectives of Swiss expert stakeholders. These include claims to data reciprocity and other reciprocal

advantages, co-authorship and collaboration opportunities, transparency on data-reuse, financial compensation, and a contribution claim based on the significance of the data contributions made by the original data collectors. While the appropriateness of some of these claims may be debatable, they should be taken seriously and discussed more openly. In order to promote health research whilst improving the quality of the health systems, these individual notions of fair data sharing deserve to be considered when promoting data sharing activities, paying particular attention to sustainable solutions that provide fair criteria for academic careers and increase high data quality and accessibility to advance science.

Abbreviations

EHR: Electronic Health Record; GDPR: General Data Protection Regulation; ICMJE: International Committee of Medical Journal Editors; NHS: National Health Service; SMAASH: Advancing SMart solutions and eliminating barriers for the Acquisition, Analysis, and Sharing of Health data in Switzerland

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12913-021-06906-2

Additional file 1. Interview guide

Acknowledgments

The authors acknowledge the financial support provided by the Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356).

Authors' contributions

LDG and AM were involved in the data collection part of the study. All authors were involved in the analysis and in defining the themes and subthemes pertinent to individual notions of fair data sharing from the perspectives of Swiss expert stakeholders. LDG wrote the first draft of the manuscript, which was then reviewed and edited by the authors. The manuscript was then finalized and approved for submission by all authors.

Funding

This study was funded by the Swiss National Science Foundation (grant number: 407440_167356). The funder played no role in the conceptualization, design, collection and analysis phases of this study, and in the writing of this manuscript.

Availability of data and materials

The dataset supporting the findings of this study is available from the corresponding author on reasonable request. The interview guide is annexed.

Declarations

Ethics approval and consent to participate

This study forms part of a bigger research project known as, SMAASH, which is exempt from requiring ethics approval according to Swiss legislation for research involving human subjects. Nonetheless, an ethics approval application for the SMAASH project was made to the cantonal ethics committee in Northwest and Central Switzerland (application number: EKNZ req-2017-00810), which confirmed the exemption of the project from requiring an ethics approval. Participants provided their verbal consent after having being informed on the objectives of the latter project, the modalities regarding the processing of their information, and their questions answered by the interviewers. Indeed, according to institutional requirements and national legislation, written informed consent was not required for participation. In this regard, all necessary regulations and guidelines for the conduct of the SMAASH project were followed.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

5.7 References

1. Hulsen T. Sharing is caring—data sharing initiatives in healthcare. Int J Environ Res Public Health. 2020;17(9):3046.
2. Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, et al. Sharing clinical trial data — a proposal from the International Committee of Medical Journal Editors. N Engl J Med. 2016;374(4):384–6.
3. Rowhani-Farid A, Allen M, Barnett AG. What incentives increase data sharing in health and medical research? A systematic review. Res Integr Peer Rev. 2017;2:4.
4. Piwowar HA, Becich MJ, Bilofsky H, Crowley RS, on behalf of the ca BIGDS, Intellectual Capital W. Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers. PLoS Med. 2008;5(9):e183..
5. Io M. Sharing clinical research data: workshop summary. Olson S, Downey AS, editors. Washington, DC: The National Academies Press; 2013. 156 p.
6. Wallis JC, Rolando E, Borgman CL. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS One. 2013;8(7):e67332.

7. Kobayashi S, Kane TB, Paton C. The privacy and security implications of open data in healthcare. Yearb Med Inform. 2018;27(1):41–7.
8. Obama B. Transparent and Open Goverment: The White House; 2009 [cited 2021 26.06.2021]. Available from: https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.
9. Data.Gov. Data Catalog: Data.Gov; 2021 [cited 2021 26.06]. Available from: https://catalog.data.gov/dataset.
10. Whitworth J. Data sharing: reaching consensus. Bull World Health Organ. 2010;88(6):468.
11. Harvard Dataverse Support. Join a growing community of Harvard and worldwide researchers who share data in the Harvard Dataverse Repository: The President and Fellows of Harvard College; 2020 [cited 2021 27.06.2021]. Available from: https://support.dataverse.harvard.edu/.
12. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Scientific Data. 2016;3(1):160018.
13. Geneviève LD, Martani A, Perneger T, Wangmo T, Elger BS. Systemic Fairness for Sharing Health Data: Perspectives From Swiss Stakeholders. Front Public Health. 2021;9:669463.
14. Geneviève LD, Martani A, Mallet MC, Wangmo T, Elger BS. Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: a systematic review. PLoS One. 2019;14(12):e0226015.
15. Popkin G. Data sharing and how it can benefit your scientific career. Nature. 2019;569(7756):445–7.
16. Institute of Medicine (US). 3, Barriers to Data Sharing. Sharing Clinical Research Data: Workshop Summary. Washington (DC): National Academies Press (US) Copyright © 2013 by the National Academy of Sciences. All rights reserved; 2013.
17. Fecher B, Friesike S, Hebing M. What drives academic data sharing? PLoS One. 2015;10(2):e0118053.
18. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data Sharing by Scientists: Practices and Perceptions. PLoS One. 2011;6(6):e21101.
19. Stuart D, Baynes G, Hrynaszkiewicz I, Allin K, Penny D, Lucraft M, et al. Practical challenges for researchers in data sharing: white paper. Springernature. 2018.

20. Teixeira da Silva JA, Dobránszki J. Do open access data files represent an academic risk? J Assoc Inf Sci Technol. 2015;66(11):2390–1..

21. Pape S, Serna-Olvera J, Tesfay WB, editors. Why open data may threaten your privacy. Workshop on Privacy and Inference; 2015.

22. Martani A, Geneviève LD, Elger B, Wangmo T. 'It's not something you can take in your hands'. Swiss experts' perspectives on health data ownership: an interview-based study. BMJ Open. 2021;11(4):e045717.

23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101.

24. Martani A, Geneviève LD, Egli SM, Erard F, Wangmo T, Elger BS. Evolution or revolution? Recommendations to Improve the Swiss Health Data Framework. Front Public Health. 2021;9:668386.

25. Kaufman A. Distributive justice, theories of. In: Chadwick R, editor. Encyclopedia of applied ethics. 2nd ed. San Diego: Academic Press; 2012. p. 842–50.

26. Lamont J, Favor C. Distributive Justice: Stanford Encyclopedia of Philosophy; 2017 [cited 2021 22.02]. Available from: https://plato.stanford.edu/entries/justice-distributive/#Strict.

27. Knoppers BM. Framework for responsible sharing of genomic and health-related data. HUGO J. 2014;8(1):3.

28. Mascalzoni D, Dove ES, Rubinstein Y, Dawkins HJS, Kole A, McCormack P, et al. International charter of principles for sharing bio-specimens and data. Eur J Hum Genet. 2015;23(6):721–8.

29. Merson L, Phong TV, Nhan LNT, Dung NT, Ngan TTD, Kinh NV, et al. Trust, respect, and reciprocity:informing culturally appropriate data-sharing practice in Vietnam. J Empirical Res Hum Res Ethics. 2015;10(3):251–63.

30. Shabani M, Obasa M. Transparency and objectivity in governance of clinical trials data sharing: current practices and approaches. Clin Trials. 2019;16(5):547–51.

31. Hopkins AM, Rowland A, Sorich MJ. Data sharing from pharmaceutical industry sponsored clinical studies: audit of data availability. BMC Med. 2018;16(1):165.

32. Boué S, Byrne M, Hayes AW, Hoeng J, Peitsch MC. Embracing transparency through data sharing. Int J Toxicol. 2018;37(6):466–71.

33. ICJME Editors. Clinical Trials 2021 [cited 2021 22.1]. Available from: http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html.

34. International Committee of Medical Journal Editors. Defining the Role of Authors and Contributors 2021 [cited 2021 25.02.2021]. Available from: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html.

35. Horton R. A fair reward. Lancet. 1998;352(9131):892.

36. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. N Engl J Med. 2017;376(17):1684–7.

37. Cole CL, Sengupta S, Rossetti S, Vawdrey DK, Halaas M, Maddox TM, et al. Ten principles for data sharing and commercialization. J Am Med Inform Assoc. 2021;28(3):646–9.

38. NHS Digital. Data Access Request Service (DARS) charges from 2020/2021 2020 [cited 2021 26.02.2021]. Available from: https://digital.nhs.uk/services/data-access-request-service-dars/data-access-request-service-dars-charges.

39. Helgesson G, Eriksson S. Authorship order. Learned Publishing 2019;32(2):106–112.

40. Brocher Foundation. Workshop: Between privacy and efficiency: the future for the health-data processing in Switzerland. A Delphi-based workshop.: Brocher Foundation; 2021 [cited 2021 1.07]. Available from: https://www.brocher.ch/en/events/443/workshop-between-privacy-and-efficiency-the-future-for-the-health-data-processing-in-switzerland-a-delphi-based-workshop/.

Chapter 6: Research Ethics in the European Influenzanet Consortium: Scoping Review

Lester Darryl Geneviève[1], Tenzin Wangmo[1], Damien Dietrich[2,3], Olivia Woolley-Meza[4], Antoine Flahault[3], Bernice Simone Elger[1,5]

[1]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

[2]Department of Radiology and Medical Informatics, Geneva University Hospitals, Geneva, Switzerland

[3]Institute of Global Health, University of Geneva, Geneva, Switzerland

[4]ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland

[5]University Center of Legal Medicine, University of Geneva, Geneva, Switzerland

Citation: Geneviève LD, Wangmo T, Dietrich D, Woolley-Meza O, Flahault A, Elger BS. Research Ethics in the European Influenzanet Consortium: Scoping Review.JMIR Public Health Surveill 2018;4(4):e67. DOI: 10.2196/publichealth.9616 [CC-BY 4.0]

“For the first time in history we can track the evolution of a pandemic in real time. Influenza viruses are notorious for their rapid mutation and unpredictable behaviour" - Margaret Chan

6.1 Abstract

Background: Influenzanet was launched in several European countries to monitor influenza-like illness during flu seasons with the help of volunteering participants and Web-based technologies. As in the case of developing fields, ethical approaches are not well developed in the collection, processing, and analysis of participants' information. Existing controversies and varying national ethical regulations can, thus, hamper efficient cross-border research collaboration to the detriment of quality disease surveillance.

Objective: This scoping review characterizes current practices on how ethical, legal, and social issues (ELSIs) pertinent to research ethics are handled by different Influenzanet country groups to analyze similarities and identify the need for further harmonization of ethical approaches.

Methods: A literature search was carried out on PubMed, Web of Science, Global Digital Library on Ethics, and Bioethics Literature Database to identify ELSIs for Influenzanet country platforms. Only English-language papers were included with publication dates from 2003 to 2017. Publications were screened for the application of bioethics principles in the implementation of country platforms. Additional publications gathered from the Influenzanet Consortium website, reference screening, and conference proceeding were screened for ELSIs.

Results: We gathered 96 papers from our search methodology. In total, 28 papers that mentioned ELSIs were identified and included in this study. The Research Ethics Committee (REC) approvals were sought for recruiting participants and collecting their data in 8 of 11 country platforms and informed e-consent was sought from participants in 9 of 11 country platforms. Furthermore, personal data protection was ensured throughout the Consortium using data anonymization before processing and analysis and using aggregated data.

Conclusions: Epidemics forecasting activities, such as Influenzanet, are beneficial; however, its benefits could be further increased through the harmonization of data gathering and ethical requirements. This objective is achievable by the Consortium. More transparency should be promoted concerning REC-approved research for Influenzanet-like systems. The validity of informed e-consent could also be increased through the provision of a user friendly and standard information sheet across the Consortium where participants agree to its terms, conditions, and privacy policies before being able to fill in the questionnaire. This will help to build trust in the general public while preventing any decline in participation.

6.2 Introduction

Web-based technologies have become an integral part of public health surveillance over the last 2 decades [1]. It is estimated that 4.3 billion people globally will have mobile broadband subscriptions by the end of 2017 [2]. Their ubiquitous availability allows volunteer citizens to engage in disease detection through digital means [3]. Real-time granular health data are, thus, collected from volunteering participants (eg, via mobile phones with global positioning), supplementing the big data collected by public health authorities and laboratories [3,4]. Combining these different data sources allows earlier and finer spatial detection of public health threats than traditional surveillance systems, permitting more appropriate preventive and mitigating measures to be deployed [4]. A successful example of such disease digital detection is the European Influenzanet Consortium.

Every year in Europe, seasonal flu brings its share of morbidity and mortality among vulnerable groups (eg, the elderly) and a rise in associated medical costs [5]. Infection with influenza virus is hard to diagnose without virological confirmation, and public health authorities usually rely on influenza-like illness (ILI) as a surveillance indicator for outbreaks [6]. This surveillance program is carried out by the European Influenza Surveillance Network (EISN), which is coordinated by the European Centre for Disease Control [7]. Since 2008, EISN has relied on ILI reports by general practitioners from national sentinels in its 30 European Union and European Economic Area countries [8]. However, this traditional surveillance system is biased by the use of nonuniform case definitions for ILI by the Member States and depends on the rate at which patients seek medical care from general practitioners, thereby reflecting only medically attended ILI incidence rates [9-11]. The general practitioner consultation rate is itself dependent on several factors that include the time delay between the onset of symptoms and health complications, the need of certificates from general practitioners for prolonged work absenteeism owing to illness, types of health insurance, and health care systems [12]. Thus, there is a non negligible underestimation of the real disease burden of influenza outbreaks [13]. The current limitations of EISN led to the development of Influenzanet, an innovative ILI surveillance network based on the active participation of public volunteers and the use of Web-based technologies to report cases of ILI, complementing data gathered by EISN [12,14].

The Influenzanet Consortium was launched in 2003 in the Netherlands and the Flemish part of Belgium [6]. Denmark, France, Ireland, Italy, Portugal, Spain, Sweden, Switzerland, and the United Kingdom have also joined this surveillance network [15]. However, very recently in 2017, the Netherlands-Belgium platforms have ceased their activities because of lack of funding, which undermines the excellent work done by these platforms in promoting the ILI

surveillance for 15 years in their countries. The platforms will resume their activities if funding is made available by May 2018. Otherwise, the platforms will terminate their activities permanently [16]. Public volunteers are usually recruited via mass media, and there are no exclusion criteria for registration (except for Sweden) [17-19]. At registration, volunteers fill in an intake questionnaire and afterwards, receive a weekly reminder by email to fill in an ILI-related symptom questionnaire [6]. Importantly, the absence of symptoms is also declared. In addition, participants are allowed to indicate symptoms of ILI for other household members in an attempt to increase the data collection for children and elderly individuals [17].

Influenzanet offers numerous advantages over EISN, including the following: ILI incidence rates are extrapolated from both medically attended and unattended patients, real-time disease-monitoring capability through the citizen participation, flexibility to changes without disturbance of the overall system functionality, uniform data collection allowing direct comparisons between countries, comparatively lower running costs, easier to increase scalability, and participant empowerment through information on prevention strategies and disease activity at local and national levels [11,13,17,20].

Nonetheless, Influenzanet also has some disadvantages, including the self-selection bias of participants (eg, underrepresentation of younger and older age groups), absence of virological confirmation of influenza cases, recruitment and motivation of participants to continuously donate their data for surveillance are problematic (eg, limited sample sizes in some countries), and limited amount and complexity of data that can be collected to ensure the continued use of platforms by participants [11,13,17].

This approach for monitoring disease involves the collection of information about users that affects their risks of influenza or complication from influenza; this information includes the demographic data, vaccination status, presence of certain medical conditions (eg, diabetes mellitus), use of food complements, daily activities, and household composition (eg, presence of children) [21]. Moreover, some national platforms have developed mobile apps as additional data collection tools; for instance, the Swiss mobile app gathers useful supplementary data through smartphone sensors (eg, inbuilt movement sensors to test for an association between the physical activity level of participants and the risk of ILI) or smartphone features (eg, test for an association between the psychological profile of participants, inferred via the list of apps installed, and attitude toward vaccination; Grippenet Switzerland, email communication, November 29, 2017). This complements the data gathered from questionnaires, allowing to answer innovative research questions and implement rational public health strategies while maintaining high privacy protection measures; for example, only highly aggregated and

summarized versions of the data are transmitted for analysis, whereas the bulk of data is stored locally on participants' smartphones. Moreover, mobile app data will not be shared with the rest of the Consortium until a framework for data sharing is set up. Nonetheless, Influenzanet data can be valuable and sensitive information. Thus, collecting such information poses ethical, legal, and social issues (ELSIs), in particular, if the collected data could be used for secondary research purposes or in the event of a cyber attack leading to data leakage.

As it is commonly the case for developing fields, ethical approaches are not yet well developed in the collection (here through Web-based technologies), processing, and analysis of participants' information. Existing controversies can, thus, be the cause of additional barriers to efficient collaboration. Furthermore, although research collaboration and comparability of data are important because epidemics do not stop at national borders, varying ethical regulations at national levels can hamper collaboration between countries. Additionally, a number of new ethical issues raised by Influenzanet-like activities do not or only partially fit traditional evaluation categories used by Research Ethics Committees (RECs) for clinical trials or data-based research. This scoping review aims to discuss ELSIs of these participatory surveillance systems. First, we characterized the current practices using findings from the literature search where we compared how issues related to research ethics are being handled by different Influenzanet country groups to analyze similarities and identify the need for further harmonization of ethical approaches. Thereafter, we carried out an ELSI analysis to suggest ways to strengthen them to lay the ground for expanding the capacity and positive impact of such systems in the future.

6.3 Methods

For this review, we followed the methodological guidance provided for conducting a scoping review [22]. Four databases, namely, PubMed, Web of Science (all databases), Global Digital Library on Ethics, and BELIT (Bioethics Literature Database) were searched to identify ELSIs for the national platforms of the Influenzanet Consortium.

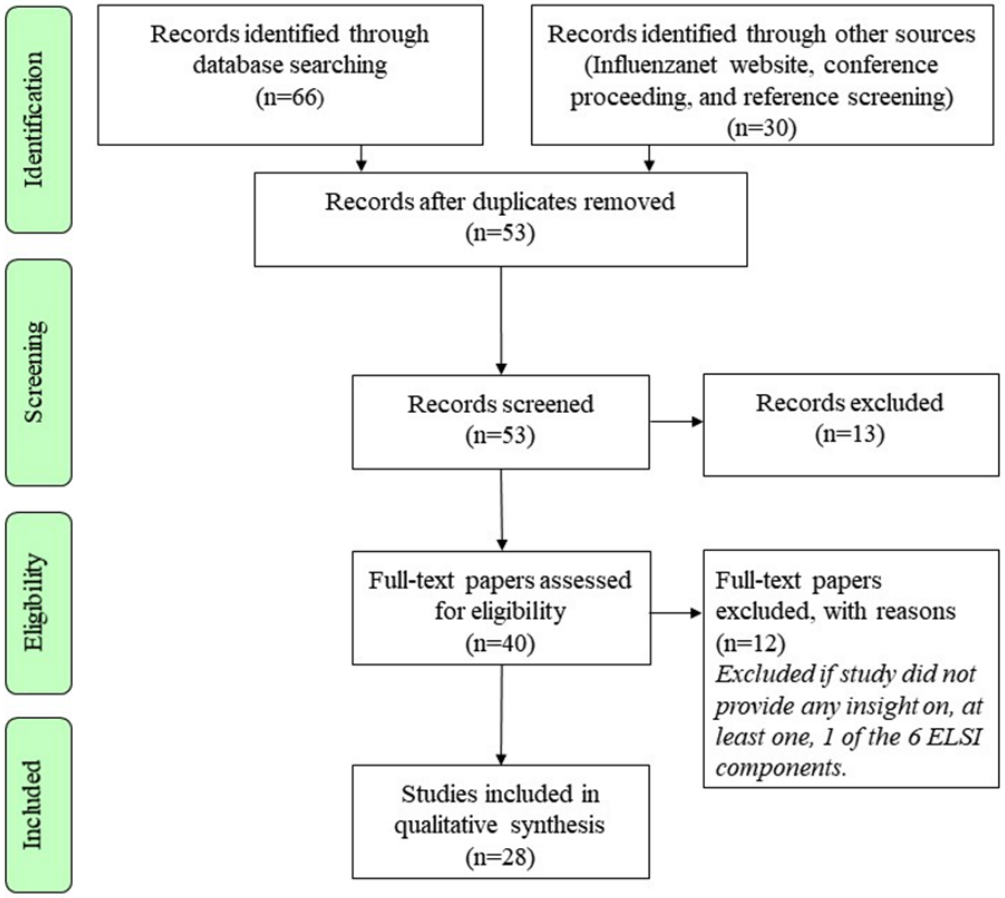
We used the following key terms: *Influenzanet*, *De Grote Griepmeting*, *Flusurvey*, *Gripenet*, *Grippenet*, *Hälsorapport*, *Influmeter*, and *Influweb*. Only English-language papers with publication dates from the last 15 years (2003-2017) were included. The search started from 2003 because the first Influenzanet national platform, *De Grote Griepmeting*, was launched in this year.

In line with the conception of modern research ethics, we searched publications related to Influenzanet for the use and application of ethical principles [23,24] in the implementation of

these national platforms; these principles are as follows: respect for autonomy (respect the decision-making capacity of Influenzanet participants through the provision of a: an informed consent and b: opt-out option); beneficence (direct and indirect benefits provided to Influenzanet participants via Web-based information on the study); nonmaleficence (prevention of informational harm to Influenzanet participants such as personal data protection measures, for example, anonymization of personal data); and justice (ensuring open and nondiscriminative participation of users to the Influenzanet network to ensure fairness in the distribution of benefits and risks) [25]. In addition, we evaluated the presence of ethical approval by an ethics committee to balance the benefits and risks to participants, future patients, or society. Additional publications found on the Influenzanet Consortium website [15] were gathered and screened for ELSIs as well. Furthermore, reference lists of included publications were searched for additional studies. Only publications mentioning at least one of these 6 ELSI components were included.

The included full-text papers were screened and analyzed independently by 2 review authors (LDG and TW) to ensure that they met the inclusion criteria of having information on the desired ELSI components previously described. Discrepancies between the 2 review authors were solved through discussion. Figure 1 illustrates the methodological process behind the selection and inclusion of publications for this review based on the PRISMA framework for systematic reviews and meta-analyses [26]. When the gathered literature did not provide sufficient information on some national platforms, country members of the Influenzanet Consortium were contacted either through email or the coordinator of the Consortium to provide additional details and to assess the veracity of the information gathered on their respective platforms.

Figure 1 The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of study selection. ELSI: ethical, legal, and social issues.



6.4 Results

The original literature search, carried out on September 11 and repeated on October 10 2017 in the databases mentioned above, identified a total of 66 potentially relevant papers (Table 1) on the Influenzanet Consortium and its national platforms (eg, *De Grote Griepmeting* for the Netherlands and Belgium, *Flusurvey* for the United Kingdom and Ireland, *Gripenet* for Portugal and Spain, *Grippenet* for France and Switzerland, *Hälsorapport* for Sweden, *Influmeter* for Denmark, and *Influweb* for Italy). No papers on Influenzanet were found on the Global Digital Library on Ethics and BELIT.

Table 1. Initial search results (date of search: September 11, 2017 and October 10, 2017).

Search terms used	Results found in PubMed (n=22), n	Results found in all Web of Science databases (n=44), n
European network: Influenzanet	10	16
National platforms		
De Grote Griepmeting	0	0
Flusurvey	3	7
Gripenet	1	1
Grippenet	6	12
Hälsorapport	0	2
Influmeter	1	1
Influweb	1	5

After the removal of duplicates (n=29), 37 papers were considered for this study. Among these 37, 4 were datasets, 5 included supporting information to other papers in the list (eg, tables and figures), 2 papers were in other languages (French and Swedish), and 2 were meeting abstracts; these 13 papers were excluded. The remaining 24 full-text papers were then screened for 6 ELSI components (informed e-consent obtained from participants; ability of participants to opt out from the study at any time; Web-based information on national platform and influenza; personal data protection measures, for example, anonymization and abiding to the national regulations on privacy, data collection, and treatment; open and nondiscriminative participation; and ethical approval by an REC or other competent entity). Only 16 of the 24 papers addressed some ethical, legal, and social components (eg, ethical approval by RECs, informed consent, etc).

Our search of the Influenzanet Consortium website [15] resulted in an additional 28 publications. After the removal of duplicates (n=14), 14 publications were considered for a detailed review of their ELSI components. Notably, 10 of the 14 papers addressed some ethical, legal, and social components.

Overall, 2 additional publications (retrieved from reference screening and a conference proceeding) were included, leading to a total of 28 papers included in our analysis, as seen in Figure 1. Table 2 reports on these 28 papers and summarizes the presence or absence of the ELSI components for each paper reviewed. However, Table 2 should be interpreted cautiously because the presence of some ELSI components does not automatically apply to all country platforms in noncountry-specific publications.

Table 2. A list of included studies (n=28) with ethical, legal, and social issue components (with ethical approval of study). All platforms listed by each paper do not satisfy the ethical, legal, and social issue components equally.

Author (year)	Platform(s) concerned	Ethical, legal, and social issue components					
		Research ethics committee	Open and non-discriminative participation	Web-based information sheet	Informed e-consent	Opt out from study	Personal data protection measures
Adler et al (2014) [27]	UK[a]	✓	✓				✓
Bajardi et al (2014) [28]	BE[b], FR[c], IT[d], NL[e], PT[f],SE[g], UK	✓	✓	✓	✓		✓
Bajardi et al (2014) [29]	BE, FR, IT, NL, PT, SE, UK	✓	✓		✓		✓
Brooks-Pollock et al (2011) [19]	UK	✓	✓	✓	✓		✓
Cantarelli et al (2014) [14]	BE, FR, IT, NL, PT, SE, UK	✓	✓	✓	✓		✓
Debin et al (2013) [30]	FR	✓		✓			✓
Debin et al (2014) [10]	FR	✓			✓		
Eames et al (2012) [31]	UK		✓				✓
Eames et al (2012) [32]	UK	✓	✓			✓	✓
Friesma et al (2009) [5]	BE, NL		✓	✓			
Guerrisi et al (2016) [33]	BE, FR, IT, NL, PT, UK		✓	✓			✓
Kjelso et al (2016) [20]	DK[h]		✓			✓	
Koppeschaar et al (2017) [17]	BE, DK, FR, IE[i], IT, NL,PT, SE, ES[j], UK	✓	✓	✓	✓		✓
Land-Zandstra et al (2016) [34]	BE, NL		✓	✓			
Loubet et al (2016) [35]	FR	✓		✓	✓	✓	✓
Loubet et al (2016) [36]	FR	✓		✓	✓	✓	✓
Marquet et al (2006) [21]	NL		✓	✓			✓
Paolotti et al (2010) [37]	IT		✓	✓			
Peppa et al (2017) [38]	UK	✓	✓		✓		
Perrotta et al (2017) [39]	IT		✓				✓
Perrotta et al (2017) [40]	IT		✓				✓
Pini et al (2017) [18]	SE					✓	✓
Smolderen et al (2007) [41]	NL		✓				✓
Tilston et al. (2010) [42]	UK	✓	✓	✓			
van Noort et al (2007) [11]	BE, NL, PT		✓	✓			
van Noort and Stollenwerk (2008) [43]	BE, IT, NL, PT		✓			✓	
van Noort et al (2015) [9]	BE, IT, NL, PT	✓		✓			

Vandendijck et al (2013) [6]	BE		✓			✓	✓

[a]UK: United Kingdom; [b]BE: Belgium; [c]FR: France; [d]IT: Italy; [e]NL: the Netherlands; [f]PT: Portugal; [g]SE: Sweden; [h]DK: Denmark; [i]IE: Ireland; [j]ES: Spain.

The Influenzanet country-specific information in Table 3 was collated from publications gathered in the literature search (Table 2) and additional information provided by country representatives. The information in Table 3 was subsequently double checked, updated, and corrected by country representatives (in many cases, authors of papers themselves) through the help of the supervisor of the Consortium who communicated our findings. This verification step was important to prevent inaccuracies resulting from the misinterpretation of the literature because the absence of an ELSI component in a publication does not automatically imply that it was not addressed by the platform(s) and temporal evolution of these platforms, where ELSI components might change over time.

Table 3 The ethical, legal, and social dimensions of Influenzanet national platforms (with research ethics approval of study).

National platform	Date of creation	Research ethics approval of study	Open and nondiscriminative participation	Web-based information sheet	Informed e-consent	Ability to opt out from the study	Personal data protection[a]
Belgium (Flanders)	2003	—[b]	✓	✓	—	✓	✓
Denmark	2013	—[c]	✓	✓	✓[c]	✓	✓
France	2012	✓	✓	✓	✓	✓	✓
Ireland	2013	✓	✓	✓	✓	✓	✓
Italy	2008	✓	✓	✓	✓	✓	✓
Portugal	2005	✓	✓	✓	✓	✓	✓
Spain	2012	✓[d]	✓	✓	✓[d]	✓	✓[e]
Sweden	2011	✓	—	✓	✓	✓	✓
Switzerland[f]	2016	✓	✓	✓	✓	✓	✓
The Netherlands	2003	—	✓	✓	—[g]	✓	✓
United Kingdom	2009	✓	✓	✓	✓	✓	✓

[a]Data anonymization before processing and analysis.

[b]Not applicable.
[c]Influmeter, email communication, August 10, 2017-September 07, 2017.
[d]GripeNet Spain, email communication, August 10, 2017.
[e]Information obtained from the Spanish national platform [44].
[f]Swiss Influenzanet platform [45]. Grippenet Switzerland, email communication, September 12, 2017.
[g]De Grote Griepmeting The Netherlands, email communication, January 10, 2018.

Only 3 of 11 national platforms (Belgium-the Netherlands and Denmark) did not seek ethical approval by REC before the launch of their platforms. The Swiss national platform [45] has obtained ethical approval for the launch of its mobile app before the start of the flu season 2017/2018. Registration and participation to the national platforms were open and nondiscriminative to all residents of the respective countries [12,14], except for the Swedish platform, where participation is through invitations only [17,18]. The Web-based information on the study was provided to all Influenzanet participants and informed electronic consents were obtained from participants in 9 of 11 national Influenzanet platforms. The electronic consent or so-called "e-consent" commonly used in studies without face-to-face contact, but where communication is entirely taking place via Web-based technologies, is another exception to, or adaptation of, traditional informed consent. Influenzanet uses this type of e-consent, where participants through a few clicks on a screen agree to the terms, conditions, and privacy policies of the research project.

The Belgium-Dutch and Danish platforms were the only exceptions where an informed e-consent was not legally required for participation (Influmeter, email communication, August 10, 2017) [6,20,21]. All Influenzanet users were allowed to withdraw from the research at any time. Participant identifiers used in Influenzanet (eg, pseudonyms and email addresses) were stored separately from the questionnaire data and not used during the data processing and analysis phases. Personal data from participants were, thus, anonymized before processing or analysis, which were performed at the aggregate level in all national platforms [14].

6.5 Discussion

Principal Findings

To the best of our knowledge, this is the first scoping review examining the similarities and differences in the implementation of research ethics for the country platforms of Influenzanet. Our comparative tables highlight the need for further clarification and harmonization of these ethical issues pertinent to citizens engaging in the digital disease surveillance across the Consortium. A number of ELSIs are similarly organized in the Consortium, for instance, participation is open and nondiscriminative to all residents of these countries (except for Sweden for representativeness and comparison purposes [17]), study information is provided to all participants, and they are free to opt out from the study. However, a number of ELSIs are

also addressed differently; for instance, REC approvals and informed e-consent were sought for recruiting participants and collecting their data in 8 of 11 and 9 of 11 platforms, respectively. The following sections of the discussion will highlight the discrepancies seen in the implementation of research ethics for different country platforms.

Overall, 8 platforms of the Consortium obtained REC approval before the start of their studies. However, it is not known how national RECs judged and approved their respective studies; for instance, they could have considered the gathered personal health-related data from participants to be fully anonymized for which no informed consent is required or considered studies to be human subject research. In the latter case, RECs would need to evaluate if the balance between potential benefits and risks for study participants is favorable and ensure that participants received adequate information on these risks and benefits. Our comparative table shows that all country platforms where REC approvals had been sought obtained informed e-consents from participants; this seems to indicate that these national RECs consider this type of citizen participatory research as human subject research and that e-consent is considered a valid form of consent in this context. The regional REC in Geneva approved the implementation of the Swiss platform and the launch of its mobile app as a data collecting tool. Considered as human subject research in Switzerland, a reader friendly informed e-consent is requested from potential participants. Because we did not have access to additional REC evaluations, further studies are needed to determine how RECs from different countries debated the ethical issues. It is also well known that national RECs, as well as RECs within the same country, may assess and balance risk-benefit ratios differently. These divergences concerning the evaluation of similar Influenzanet projects in different countries could interfere with the harmonization of ethical approaches. Thus, we suggest more transparency in terms of ethical issues related to this type of technology-driven public health research. For instance, project leaders of national Influenzanet platforms could publish a summary of how RECs evaluated and debated the ethical issues of their respective studies (eg, if their studies fall under the category of human subject research and, thereby, need informed consent procedures, etc). Such transparency could help to harmonize the ethical approaches to be adopted by the country platforms even further.

REC approvals were not sought for the Belgian-Dutch (*De Grote Griepmeting*) and Danish (*Influmeter*) platforms (Influmeter, email communication, August 10, 2017) [6,20,21]. However, their studies abided by their national legislation on privacy and personal data protection (Influmeter, email communication, August 10, 2017) [6,20,21,46]; for instance, the *De Grote Griepmeting* privacy regulation was approved by the Dutch Data Protection Authority [21]. According to the Belgian and Dutch legislations, these are observational studies because

no physical or psychological intervention is intended on participants [6,21,47]. Concerning *Influmeter*, the Danish platform is exempted from the REC approval for the following reasons: the Danish Data Protection Agency does not consider emails exchanged between study participants and *Influmeter* to be sensitive personal information; there is an automatic implicit consent from study participants because of its voluntary nature even if sensitive information is gathered (eg, health and coarse-grained geographical data); and the data manager of *Influmeter*, *Statens Serum Institute* that hosts a large proportion of Danish health data [48] received a broad permission from the Danish Data Protection Agency (record number: 2008-54-0474), which covers the surveillance of infectious diseases and identifiable sensitive information gathered by *Influmeter* (Influmeter, email communication, September 7, 2017) [20].

Another ethical issue arises in Influenzanet owing to the ability of participants to record personal data on other household members (eg, the elderly persons and children). Gathering data on underrepresented age groups is important, specifically, when they are the ones most vulnerable to influenza in terms of morbidity and mortality [5]. However, it is difficult to verify whether these family members, in particular, are legally competent and could provide consent themselves, having expressed their will for their personal information to be recorded by the participating family member. It would be interesting to evaluate through future research whether RECs have considered this issue or have simply considered the data collected from other family members to be anonymous and, thus, not identifiable.

Informed e-consent was gathered from study participants from all platforms with the only exceptions being the Belgian and Dutch platforms, which are mirror websites of each other (*De Grote Griepmeting*, email communication, January 10, 2018). It can be argued that there is an automatic implicit consent for Belgian and Dutch participants because registration to the study is voluntary. The Belgium-Dutch platforms [16] might consider providing an informed e-consent option to their participants in an attempt to harmonize consent practices across the Consortium if ever they resume their activities in the future. However, we noted that it is not clear whether informed consent was legally necessary in Belgium ("Law on experiments involving the human subject" of May 2004) and the Netherlands because of the observational natures of their studies (*De Grote Griepmeting*, the Netherlands, email communication, January 10, 2018) [6,21,47]. Indeed, the Belgian and Dutch legislations acknowledge the need for informed consent for interventional studies because of the potential physical or psychological harm to participants [6,21,47], but the legislations do not clearly define how broad the category of observational studies is; for instance, a detailed questionnaire revealing some very personal information can be seen as an intervention in the Netherlands [49]. Nonetheless, it is important

to understand the shift from typical physical or psychological harm seen in medical research to informational harm in public health research involving Web-based communities of volunteer citizens or big data (eg, data discrimination) with the latter having potential repercussions on the physical and mental states of study participants (eg, stigmatization and discrimination for health insurance coverage) [50,51].

Critics might say that e-consent gathered from Influenzanet participants does not represent valid informed consent because researchers cannot control that participants read and understood the information. Although information on studies is available on their respective websites, participants usually have to look at different sections of the website to gather pertinent information on the study (eg, its goals, privacy policies, etc), which is tedious and unlikely to be read in detail. The Swiss legislation requires researchers to ensure that participants have understood all the information (provided in the written information form) through personal contact. According to the *Ethics Guidelines for Internet-Mediated Research* by the British Psychological Society, a valid consent can be assumed if there is an information sheet defining the study objectives and exact nature of questions before filling in the questionnaire, including a check box at its beginning and end where participants can tick in to give their explicit consent [52]. In addition, the Society recommends using a proper wording for "I agree" statements to encourage participants to read the information sheet, which should also include their rights to withdraw from the study at any time in a user friendly manner [52]. The Consortium could follow these guidelines to enhance its e-consent procedures. It is important to ensure that participants agreed (by clicking on an "I agree" checkbox) to the terms, conditions, and privacy policies listed on the informed consent sheet before being allowed to fill in the questionnaire. Such measures should be taken to ensure the validity of the informed consent from participants and for harmonization purposes, a standard information sheet could be used throughout the Consortium.

Personal data protection was ensured throughout the Consortium because participant data are pseudonymized, that is, participants' personal identifiers are replaced with pseudonyms. Furthermore, any data that are shared with the public through the Web portal are fully anonymized and highly aggregated [14]. As to the anonymized data shared with members of the Consortium, it can be aggregated or not depending on whether the national partner who owns the data agrees to the request. It is also worth noting that sharing of Influenzanet data with internal and external researchers should not pose *a priori* any legal barriers because the General Data Protection Regulation of the European Parliament and the Council will not apply to data being rendered anonymous, that is, the information is not linked to an identified or identifiable

natural person (Recital 26) and shared for research and statistical purposes [53]. From information gathered on some Influenzanet platforms (ie, France and Switzerland), it appears that linked anonymization [54] is involved in ensuring data protection, whereby participants' email addresses and pseudonyms are stripped from the gathered health-related data before analysis and stored separately. It will be difficult to conclude how challenging or easy it could be to reidentify participants directly or indirectly from specific combination of variables (which are anonymous if considered in isolation but permit the identification if several of them are combined), in particular for vulnerable groups where privacy risks are higher [48] or groups with rare variable entries, for instance, large household compositions with >6 persons (that only account for 2% of households in the European Union [55]).

Our results show that there is a need to harmonize consent requirements and practice throughout the Consortium because differences in the abovementioned national consent requirements could be the source of obstacles for the next generation of data collection, which may include collection and sharing of more sensitive data on a larger scale. Thus, these differences could hinder future productive cross-national collaboration, which will be detrimental to research and quality disease surveillance. The consent requirements and practice could be harmonized through international or European Union regulations. However, this could take some time; for instance, General Data Protection Regulation aims to harmonize data protection laws within the 28 European Union Member States to ensure an equivalent level of protection and freedoms of individuals within the European Union. At the same time, it will protect cross-border flows of personal data on European Union citizens to international organizations or third countries [56,57]. A quicker harmonization alternative would be for all national platforms of Influenzanet to use the strictest consent requirements and practice currently used by one of their platforms to find common ground. This particular platform would then be used as a benchmark for other country platforms to harmonize their practices.

Limitations

Only English-language papers were screened for this study for practical reasons. It is, thus, possible that pertinent papers in other languages were omitted but they could have provided better insight into how issues related to research ethics are being handled by the Influenzanet Consortium. In addition, most of our conclusions are based on the gathered literature, and our interpretations might be biased by the incomplete reporting of all ELSI components in some publications as a matter of space or pertinance to their study objectives. Consequently, despite our collaboration with the Consortium to verify our claims, we cannot guarantee that our interpretations are error free. Another limitation is that we do not have access to projects'

evaluation reports from RECs (except for Switzerland) to understand how they judge and approve such research projects. This would have been beneficial to the understanding of national differences in project evaluation by RECs, which could then serve as the basis for further harmonization of these ethical approaches.

6.6 Conclusions

Epidemics forecasting activities such as Influenzanet are beneficial. Harmonized criteria for dealing with ethical issues are urgently needed internationally to ensure comparability of data and maximize participants' trust. Approaches used in handling issues related to research ethics by different Influenzanet platforms seem to be similar in many, although not all, ELSIs at present. Thus, harmonizing ethical requirements across the Consortium is feasible and could be achieved through the adoption of the strictest ethical requirements and practice currently used by one platform across the Consortium. Nonetheless, despite being similar, it does not automatically mean that these ethical approaches are adequately regulated. We recommend more transparency in terms of ethical issues related to this type of technology-driven public health research. These transparency modifications related to the current ELSIs of Influenzanet will help to build trust among the members of the general public, in particular, if they are properly informed about the expected benefits and potential risks their participation entails; this will prevent any decline in participation, which might be triggered by mediatization, including exaggeration, of the risks of public health surveillance using Web-based communities of volunteer citizens.

Moreover, this type of research has the potential to save many lives in the future because it has proven through its flexibility, easy implementation, and adaptability to different countries' requirements for data collection to serve as a potentially effective and relatively low-cost surveillance tool for other diseases of public health importance (eg, Middle East respiratory syndrome or Ebola) [12,17]. These characteristics could allow Influenzanet to be deployed in low- and middle-income countries to monitor emerging and reemerging infectious diseases [17].

Our suggested harmonization measures and approach for data gathering and ethical requirements for Influenzanet apply to other Influenzanet-like systems. Moreover, we also suggest that such systems increase the validity of their informed e-consent procedures by following the excellent ethical guidelines provided by the British Psychological Society. Such measures will further increase the benefits Influenzanet and Influenzanet-like systems could bring to society by promoting the comparability of data and safeguarding participants' trust on

which they rely almost completely for data collection. These will ensure that Influenzanet-like systems are making even greater substantial contributions to global public health and reduce the health inequalities in societies worldwide through better targeting of public health interventions in line with the concept of *precision global health* in the digital age [58].

Acknowledgments

We sincerely thank the coordinator of the Influenzanet Consortium, Daniela Paolotti, and country representatives for verifying the veracity of our findings and for providing updated additional information on their respective national platforms. OW-M, DD, and AF thank Dirk Helbing, Paul Lukowicz, Daniela Paolotti, Lloyd Sanders, and Marco Hirsch for the guidance and collaboration on Grippenet (www.grippenet.ch), the Swiss member of Influenzanet (www.influenzanet.eu). LDG also thanks Andrea Martani and Christophe Schneble. LDG, TW, and BSE acknowledge the financial support provided by the Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356 and SNF NRP-75 big data, grant number 407540_167211). The views expressed in this article are those of the authors and not necessarily those of the funder. This work was partially funded by the European Community's H2020 Program under the funding scheme "FETPROACT-1-2014: Global Systems Science (GSS)'," grant agreement 641191 "CIMPLEX: Bringing CItizens, Models and Data together in Participatory, Interactive SociaL EXploratories".

Authors' Contributions

All 6 authors contributed to the conception of this paper. LDG undertook the initial scoping literature search and was assisted by TW and BSE. TW and BSE supported LDG on the synthesis of the methodology and results sections. All authors contributed to the writing, editing, and critical evaluation of the manuscript. They approved the submission of the final version of the manuscript.

Conflicts of Interest

AF is one of the project leaders, and OW-M and DD are project coordinators of the Swiss Influenzanet platform.

Abbreviations

BELIT: Bioethics Literature Database

EISN: European Influenza Surveillance Network

ELSI: ethical, legal, and social issues

ILI: influenza-like illness

REC: Research Ethics Committee

6.7 References

1. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
2. ICT FactsFigures 2017. 2017 Jul. ITU. URL:http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017. pdf [accessed 2017-12-08] [WebCite Cache ID 6vYQzyEgK]
3. Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. N Engl J Med 2013 Aug 1;369(5):401-404. [doi: 10.1056/NEJMp1307752] [Medline: 23822655]
4. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big Data for Infectious Disease Surveillance and Modeling. J Infect Dis 2016 Dec 01;214(suppl_4):S375-S379 [FREE Full text] [doi: 10.1093/infdis/jiw400] [Medline: 28830113]
5. Friesema IHM, Koppeschaar CE, Donker GA, Dijkstra F, van Noort SP, Smallenburg R, et al. Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands. Vaccine 2009 Oct 23;27(45):6353-6357. [doi: 10.1016/j.vaccine.2009.05.042] [Medline: 19840672]
6. Vandendijck Y, Faes C, Hens N. Eight years of the Great Influenza Survey to monitor influenza-like illness in Flanders. PLoS One 2013;8(5) [FREE Full text] [doi: 10.1371/journal.pone.0064156] [Medline: 23691162]
7. European Centre for Disease Prevention and Control. European Influenza Surveillance Network (EISN) URL:https://ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/eisn [accessed 2017-12-09] [WebCite Cache ID 6va45FX5Y]
8. Snacken R, Brown C. New developments of influenza surveillance in Europe. Euro Surveill 2015 Jan 29;20(4) [FREE Full text] [Medline: 25655056]
9. van Noort SP, Codeço CT, Koppeschaar CE, van Ranst M, Paolotti D, Gomes MGM. Ten-year performance of Influenzanet: ILI time series, risks, vaccine effects, and care-seeking behaviour. Epidemics 2015 Dec;13:28-36 [FREE Full text] [doi:10.1016/j.epidem.2015.05.001] [Medline: 26616039]
10. Debin M, Colizza V, Blanchon T, Hanslik T, Turbelin C, Falchi A. Effectiveness of 2012-2013 influenza vaccine against influenza-like illness in general population: estimation in a French web-based cohort. Hum Vaccin Immunother 2014;10(3):536-543 [FREE Full text] [Medline: 24343049]

11. van Noort SP, Muehlen M, Rebelo de Andrade H, Koppeschaar C, Lima Lourenço JM, Gomes MGM. Gripenet: an internet-based system to monitor influenza-like illness uniformly across Europe Internet. Eurosurveillance 2007 [FREE Full text] [doi: 10.2807/esm.12.07.00722-en]
12. Paolotti D, Carnahan A, Colizza V, Eames K, Edmunds J, Gomes G, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. Clin Microbiol Infect 2014 Jan;20(1):17-21 [FREE Full text] [doi: 10.1111/1469-0691.12477] [Medline: 24350723]
13. Wójcik OP, Brownstein JS, Chunara R, Johansson MA. Public health for the people: participatory infectious disease surveillance in the digital age. Emerg Themes Epidemiol 2014;11:7 [FREE Full text] [doi: 10.1186/1742-7622-11-7] [Medline: 24991229]
14. Cantarelli P, Debin M, Turbelin C, Poletto C, Blanchon T, Falchi A, et al. The representativeness of a European multi-center network for influenza-like-illness participatory surveillance. BMC Public Health 2014 Sep 20;14:984 [FREE Full text] [doi: 10.1186/1471-2458-14-984] [Medline: 25240865]
15. Influenzanet. Influenzanet: a network of European citizens fighting against influenza URL:https://www.influenzanet.eu/ en/ [accessed 2017-12-09] [WebCite Cache ID 6va4T6dwI]
16. De Grote GriepMeting. URL:https://www.degrotegriepmeting.be [accessed 2017-12-08] [WebCite Cache ID 6vYZj6dRG]
17. Koppeschaar CE, Colizza V, Guerrisi C, Turbelin C, Duggan J, Edmunds WJ, et al. Influenzanet: Citizens Among 10 Countries Collaborating to Monitor Influenza in Europe. JMIR Public Health Surveill 2017 Dec 19;3(3):e66 [FREE Full text] [doi: 10.2196/publichealth.7429] [Medline: 28928112]
18. Pini A, Merk H, Carnahan A, Galanis I, van Straten E, Danis K, et al. High added value of a population-based participatory surveillance system for community acute gastrointestinal, respiratory and influenza-like illnesses in Sweden, 2013-2014 using the web. Epidemiol Infect 2017 Apr;145(6):1193-1202. [doi: 10.1017/S0950268816003290] [Medline: 28137317]
19. Brooks-Pollock E, Tilston N, Edmunds WJ, Eames KTD. Using an online survey of healthcare-seeking behaviour to estimate the magnitude and severity of the 2009 H1N1v influenza epidemic in England. BMC Infect Dis 2011;11:68 [FREE Full text] [doi: 10.1186/1471-2334-11-68] [Medline: 21410965]

20. Kjelsø C, Galle M, Bang H, Ethelberg S, Krause TG. Influmeter - an online tool for self-reporting of influenza-like illness in Denmark. Infect Dis (Lond) 2016 Apr;48(4):322-327. [doi: 10.3109/23744235.2015.1122224] [Medline: 26654752]
21. Marquet RL, Bartelds AIM, van Noort SP, Koppeschaar CE, Paget J, Schellevis FG, et al. Internet-based monitoring of influenza-like illness (ILI) in the general population of the Netherlands during the 2003-2004 influenza season. BMC Public Health 2006 Oct 04;6:242 [FREE Full text] [doi: 10.1186/1471-2458-6-242] [Medline: 17018161]
22. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
23. Beauchamp TL, Childress JF. Principles of biomedical ethics, 5th ed. NY: Oxford University Press; 2001.
24. Ryan KJ, Brady JV, Cooke RE, Height DI, Jonsen AR, King P, et al. Ethical Principles and Guidelines for the Protection of Human Subjects of Research. In: The Belmont Report. Washington D.C. 20402: U.S. Gov. Print. Off; Sep 30, 1978.
25. Beauchamp TL. Methods and principles in biomedical ethics. J Med Ethics 2003 Oct;29(5):269-274 [FREE Full text] [Medline: 14519835]
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Adler AJ, Eames KTD, Funk S, Edmunds WJ. Incidence and risk factors for influenza-like-illness in the UK: online surveillance using Flusurvey. BMC Infect Dis 2014 May 01;14:232 [FREE Full text] [doi: 10.1186/1471-2334-14-232] [Medline: 24885043]
28. Bajardi P, Paolotti D, Vespignani A, Eames K, Funk S, Edmunds WJ, et al. Association between recruitment methods and attrition in Internet-based studies. PLoS One 2014;9(12):e114925 [FREE Full text] [doi: 10.1371/journal.pone.0114925] [Medline: 25490045]
29. Bajardi P, Vespignani A, Funk S, Eames KT, Edmunds WJ, Turbelin C, et al. Determinants of follow-up participation in the Internet-based European influenza surveillance platform Influenzanet. J Med Internet Res 2014;16(3):e78 [FREE Full text] [doi: 10.2196/jmir.3010] [Medline: 24613818]
30. Debin M, Turbelin C, Blanchon T, Bonmarin I, Falchi A, Hanslik T, et al. Evaluating the feasibility and participants' representativeness of an online nationwide surveillance system

for influenza in France. PLoS One 2013;8(9):e73675 [FREE Full text] [doi: 10.1371/journal.pone.0073675] [Medline: 24040020]

31. Eames KTD, Brooks-Pollock E, Paolotti D, Perosa M, Gioannini C, Edmunds WJ. Rapid assessment of influenza vaccine effectiveness: analysis of an internet-based cohort. Epidemiol Infect 2012 Jul;140(7):1309-1315. [doi:10.1017/S0950268811001804] [Medline: 21906412]
32. Eames KTD, Tilston NL, Brooks-Pollock E, Edmunds WJ. Measured dynamic social contact patterns explain the spread of H1N1v influenza. PLoS Comput Biol 2012;8(3):e1002425 [FREE Full text] [doi: 10.1371/journal.pcbi.1002425] [Medline: 22412366]
33. Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, et al. Participatory Syndromic Surveillance of Influenza in Europe. J Infect Dis 2016 Dec 01;214(suppl_4):S386-S392. [doi: 10.1093/infdis/jiw280] [Medline: 28830105]
34. Land-Zanstra A, van Beusekom M, Koppeschaar K, van den Broek J. Motivation and learning impact of Dutch flu-trackers. Journal of Science Communication 2016;15(01):1-26.
35. Loubet P, Guerrisi C, Turbelin C, Blondel B, Launay O, Bardou M, et al. First nationwide web-based surveillance system for influenza-like illness in pregnant women: participation and representativeness of the French G-GrippeNet cohort. BMC Public Health 2016 Mar 11;16:253 [FREE Full text] [doi: 10.1186/s12889-016-2899-y] [Medline: 26969654]
36. Loubet P, Guerrisi C, Turbelin C, Blondel B, Launay O, Bardou M, et al. Influenza during pregnancy: Incidence, vaccination coverage and attitudes toward vaccination in the French web-based cohort G-GrippeNet. Vaccine 2016 Apr 29;34(20):2390-2396. [doi: 10.1016/j.vaccine.2016.03.034] [Medline: 27013430]
37. Paolotti D, Gioannini C, Colizza V, Vespignani A. Internet-based monitoring system for influenza-like illness: H1N1 surveillance in Italy. 2010 Presented at: 3rd International ICST Conference on Electronic Healthcare for the 21st century; 13.12.2010 - 15.12.2010; Casablanca, Morocco URL:https://www.isi.it/wp_blobs/.../ influweb-ehealth2010_submitted_1384514049.pdf
38. Peppa M, Edmunds WJ, Funk S. Disease severity determines health-seeking behaviour amongst individuals with influenza-like illness in an internet-based cohort. BMC Infect Dis 2017 Mar 31;17(1):238 [FREE Full text] [doi:10.1186/s12879-017-2337-5] [Medline: 28359335]

39. Perrotta D, Bella A, Rizzo C, Paolotti D. Participatory Online Surveillance as a Supplementary Tool to Sentinel Doctors for Influenza-Like Illness Surveillance in Italy. PLoS One 2017;12(1):e0169801 [FREE Full text] [doi:10.1371/journal.pone.0169801] [Medline: 28076411]
40. Perrotta D, Tizzoni M, Paolotti D. Proceedings of the 26th International Conference on World Wide Web. In; 2017. Using Participatory Web-based Surveillance Data to Improve Seasonal Influenza Forecasting in Italy URL:https://doi.org/10.1145/ 3038912.3052670 [accessed 2018-09-10] [WebCite Cache ID 72KSK39sh]
41. Smolderen KGE, Vingerhoets AJJM, Croon MA, Denollet J. Personality, psychological stress, and self-reported influenza symptomatology. BMC Public Health 2007 Nov 23;7:339 [FREE Full text] [doi: 10.1186/1471-2458-7-339] [Medline: 18036207]
42. Tilston NL, Eames KTD, Paolotti D, Ealden T, Edmunds WJ. Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic. BMC Public Health 2010 Oct 27;10:650 [FREE Full text] [doi:10.1186/1471-2458-10-650] [Medline: 20979640]
43. van Noort S, Stollenwerk N. From dynamical processes to likelihood functions: an epidemiological application to influenza. 2008 Presented at: International Conference on Computational and Mathematical Methods in Science and Engineering; 2008; Murcia, Spain p. a-661.
44. GripeNet.es. URL:https://www.gripenet.es/ [accessed 2017-12-08] [WebCite Cache ID 6vYSX5uHD]
45. Grippenet.ch. URL:http://www.grippenet.ch/ [accessed 2017-12-08] [WebCite Cache ID 6vYS2omSa]
46. Hoeyer K. Denmark at a Crossroad? Intensified Data Sourcing in a Research Radical Country. In: Mittelstadt BD, Floridi L, editors. The Ethics of Biomedical Big Data. Switzerland: Springer International Publishing AG; 2016:73-93.
47. Service public Fédéral Justice. Loi relative aux expérimentations sur la personne humaine URL:http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=2004050732&table_name=loi [accessed 2017-10-11] [WebCite Cache ID 6va53sJgB]
48. Danish Council of Ethics. 2015. Research with health data and biological material in Denmark URL:http://www.etiskraad.dk/ ~/media/Etisk-Raad/en/Publications/Research-with-health-data-and-biological-material-in-Denmark-Statement-2015. pdf?la=da [accessed 2017-12-09] [WebCite Cache ID 6va5FW4bI]

49. Ramirez I. Navigating the maze of requirements for obtaining approval of non-interventional studies (NIS) in the European Union. Ger Med Sci 2015;13:Doc21 [FREE Full text] [doi: 10.3205/000225] [Medline: 26633964]
50. Metcalf J, Crawford K. Where are human subjects in Big Data research? The emerging ethics divide. Big Data & Society 2016 Jun 1;3(1):205395171665021. [doi: 10.1177/2053951716650211]
51. Fairfield J, Shtein H. Big Data, Big Problems: Emerging Issues in the Ethics of Data Science and Journalism. Journal of Mass Media Ethics 2014 Jan 16;29(1):38-51. [doi: 10.1080/08900523.2014.863126]
52. Hewson C, Buchanan T, Brown I, Coulson N, Hagger-Johnson G, Joinson A, et al. Ethics Guidelines for Internet-Mediated Research. Leicester: The British Psychological Society; 2017 Apr. URL:https://beta.bps.org.uk/news-and-policy/ ethics-guidelines-internet-mediated-research-2017 [accessed 2017-10-30] [WebCite Cache ID 6vYTFleqY]
53. Intersoft Consulting. 2018. Recital 26 - Not applicable to anonymous data URL:https://gdpr-info.eu/recitals/no-26/ [accessed 2018-06-04] [WebCite Cache ID 6zveqKDMo]
54. Elger BS, Caplan AL. Consent and anonymization in research involving biobanks: differing terms and norms present serious barriers to an international framework. EMBO Rep 2006 Jul;7(7):661-666 [FREE Full text] [doi: 10.1038/sj.embor.7400740] [Medline: 16819458]
55. Eurostat. People in the EU - statistics on household and family structures URL:http://ec.europa.eu/eurostat/statistics-explained/ index.php/People_in_the_EU_%E2%80%93_statistics_on_household_and_family_structures [accessed 2017-12-08] [WebCite Cache ID 6vYZAy7nZ]
56. Preite F, Salardi S, Gesuita R, Villani S, Trerotoli P, Guardabasso V, et al. The new European regulation on personal data protection: significant aspects for data processing for scientific research purposes. EBPH 2017;4(2) [FREE Full text] [doi:10.2427/12286]
57. Mostert M, Bredenoord AL, Biesaart MCIH, van Delden JJM. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet 2016 Dec;24(7):956-960 [FREE Full text] [doi:10.1038/ejhg.2015.239] [Medline: 26554881]
58. Flahault A, Geissbuhler A, Guessous I, Guérin P, Bolon I, Salathé M, et al. Precision global health in the digital age. Swiss Med Wkly 2017;147 [FREE Full text] [doi: 10.4414/smw.2017.14423] [Medline: 28421566]

Chapter 7: Participatory Disease Surveillance Systems: Ethical Framework

Lester Darryl Geneviève[1], Andrea Martani[1], Tenzin Wangmo[1], Daniela Paolotti[2], Carl Koppeschaar[3], Charlotte Kjelsø[4], Caroline Guerrisi[5], Marco Hirsch[6], Olivia Woolley-Meza[7,8], Paul Lukowicz[6], Antoine Flahault[9], Bernice Simone Elger[1,10]

[1]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

[2] Institute for Scientific Interchange Foundation, Torino, Italy

[3] De Grote Griepmeting, Science in Action BV, Amsterdam, Netherlands

[4] Statens Serum Institut, København, Denmark

[5] Sorbonne Université, Institut National de la Santé et de la Recherche Médicale, Institut Pierre Louis d'Épidémiologie et de Santé Publique, Paris, France

[6] German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany

[7] ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland

[8] Novartis Pharma AG, Basel, Switzerland

[9] Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[10] University Center of Legal Medicine, University of Geneva, Geneva, Switzerland

Citation: Geneviève LD, Martani A, Wangmo T, Paolotti D, Koppeschaar C, Kjelsø C, Guerrisi C, Hirsch M, Woolley-Meza O, Lukowicz P, Flahault A, Elger BS. Participatory Disease Surveillance Systems: Ethical Framework. J Med Internet Res 2019;21(5):e12273. DOI: 10.2196/12273 [CC BY 4.0]

"Strong health and disease surveillance systems halt epidemics that take lives, disrupt economies, and pose global health security threats." - Tedros Adhanom

7.1 Abstract

Advances in information technology are changing public health at an unprecedented rate. Participatory surveillance systems are contributing to public health by actively engaging digital (eg, Web-based) communities of volunteer citizens to report symptoms and other pertinent information on public health threats and also by empowering individuals to promptly respond to them. However, this digital model raises ethical issues on top of those inherent in traditional forms of public health surveillance. Research ethics are undergoing significant changes in the digital era where not only participants' physical and psychological well-being but also the protection of their sensitive data have to be considered. In this paper, the digital platform of Influenzanet is used as a case study to illustrate those ethical challenges posed to participatory surveillance systems using digital platforms and mobile apps. These ethical challenges include the implementation of electronic consent, the protection of participants' privacy, the promotion of justice, and the need for interdisciplinary capacity building of research ethics committees. On the basis of our analysis, we propose a framework to regulate and strengthen ethical approaches in the field of digital public health surveillance.

7.2 Introduction

Advances in information technology are changing medical research [1] and public health at an unprecedented rate [2]. One of the most evident changes is that it has become easy for members of the general public to contribute to public health surveillance, practice, and policy [2] by sharing personal and health-related information through digital media. The pervasiveness of technology is underscored by the fact that as of 2018, almost 4 billion people are estimated to have access to the internet [3], and there are over 318,000 health-related mobile apps [4]. In public health surveillance (ie, public health data collection and analysis to inform public health practice [5]), vast real-time health data from informal sources (eg, health-related mobile apps and twitter) allow an early detection, prevention, and monitoring of public health threats and the potential for a prompt response from authorities to mitigate them. These informal sources have facilitated the reporting of diseases by complementing and reducing the time information is transmitted in multilevel public health infrastructures. Consequently, around the world, several early warning systems are now using this innovative approach [6].

Such activities have been termed *digital epidemiology* or *digital disease detection* [7,8]. Digital epidemiology can either be performed for nonresearch or research purposes. On the one hand, if the aims of the surveillance system are simply to monitor, control, and respond to health threats by producing data on the affected population, then this serves nonresearch purposes. On the other hand, if the purpose of the surveillance system is either to contribute to or to produce

generalizable knowledge, potentially applicable to different populations and settings, then it serves research purposes [9].

Data in digital epidemiology can be obtained through 2 distinct approaches, with similar public health objectives but usually different challenges [10]. With a passive approach for data collection, data subjects are not directly informed that their everyday data (stored, eg, on social platforms, blogs, and Web search queries) are being mined and processed by advanced algorithms involved in big data analytics to monitor or predict disease outbreaks [10-12]. One of the first notable examples of such passive data collection approach to digital epidemiology was *Google Flu Trends* (GFT). There are numerous challenges to this approach, including *big data hubris* and unstable algorithm dynamics. *Big data hubris* states that big data are simply a replacement for data collected and analyzed by conventional means rather than an adjunct to traditional public health surveillance. Unstable algorithm dynamics refer to the continuous changes made by the company to the search algorithms, as a means of improving their searching capabilities by incorporating, for instance, new search terms. However, as the case of GFT demonstrated, these improvements led to biased estimates [13]. Nonetheless, *big data* surveillance offers an unprecedented opportunity to monitor in a timely manner the spatial and temporal evolutions of epidemics with increased granularity compared with more traditional surveillance systems, provided that the potential flaws of *big data analytics* are taken into consideration [12,13].

On the other hand, a *participatory disease surveillance* system has an active approach involving digital communities of volunteer citizens who consciously provide data. This can be done either interactively by reporting their symptoms and other relevant information through an appropriate interface or by donating sensor data (eg, location traces) from their digital devices. Such a participatory approach not only supports the detection of potential public health threats but also empowers individuals to reflect on them and adapt their behavior accordingly [10]. An example of such participatory disease surveillance system is the European Influenzanet Consortium, which monitors *influenza-like illness* (ILI) activity during flu seasons, with data from volunteer citizens using digital national platforms and in some cases, mobile apps [14,15]. Details on the Influenzanet Consortium and its inclusion as an adequate model to illustrate ethical issues pertinent to participatory disease surveillance systems have been covered in a previous publication [16].

In 2009, the Influenzanet Consortium was formed to standardize practices among the individual national *ILI* surveillance platforms to promote collaboration [17]. Recent research started

exploring the use of crowdsourcing for detection of epidemic flu spreading, improving the self-reporting experience of symptoms with a more user-friendly mobile phone apps, and enriching the data with context data recorded by the phone's sensors [15]. The Consortium follows the top-down model of citizen participation [2], which guarantees the scientific requirements and integrity of the disease surveillance network while relying on volunteer citizens' data. This technology-driven public health surveillance has some benefits for its participants. Real-time information on *ILI* activity at local and national levels is provided to participants, who are also advised on strategies for disease prevention [17,18]. Importantly, the participants contribute directly to the ultimate goal of this public health initiative by providing real-time granular health-related data on *ILI* [19]. Such information complements the data of the *European Influenza Surveillance Network* (EISN) at finer levels, as EISN receives mainly epidemiological and virological data from its network of general practitioners [17,18]. The large cohorts of Influenzanet (eg, over 36,000 volunteer citizens for the flu season 2015/2016) also allow detection of even small epidemics of *ILI* [17,20]. This early detection [21] could potentially enable timely mitigation strategies to reduce the health burden of influenza and decrease health expenditures associated with increased hospitalization and treatment. In addition, Influenzanet enables research and the study of subgroups, for example, influenza vaccine effectiveness in vaccinated groups [22], attitudes toward vaccination [23], health status of population outside the health care system [24], and differences in medical care–seeking behavior across the European Union (EU) [20].

However, such top-down model of citizen participation in surveillance, specifically for research purposes, raises its own set of ethical issues on top of those inherent in traditional forms. Participants are involved actively in scientific research [2], but researchers have limited personal interaction with participants to ensure that they have indeed understood the research information provided on the national platforms or mobile apps and potential risks that their participation entails. In addition, participants could be influenced by the promise of expected benefits and the imperative of altruism. Therefore, it becomes a challenge to combine protection of research participants with the promotion of high-quality data collection for ethically acceptable research purposes [25]. In participatory disease surveillance systems, these are closely linked and mutually dependent on each other. On the one hand, ensuring participants' trust and engagement through adequate safeguards is crucial for the sustainability and quality of these surveillance systems. If participants perceive the risk of privacy violation, they might refrain from giving important information, thus, affecting the effectiveness of disease surveillance and subsequent future public health interventions [26]. On the other hand, if

surveillance systems follow low-quality standards and operate outside an ethical framework, protection and collaborations of data subjects cannot be secured. Numerous ethical frameworks have been developed in the field of public health surveillance and the use of big data [27-29]. However, a 2017 systematic review on ethical issues of public health surveillance revealed that there is a need for more context-specific analyses to guide public health practice [5]. Consequently, providing an ethical framework for the regulation of such innovative participatory surveillance methods, using a real-world example, becomes of utmost importance.

In this paper, we use Influenzanet to illustrate challenges in protection of health and other sensitive information reported in participatory disease surveillance systems. We discuss and analyze challenges and needs of participant consent in surveillance and research using participant surveillance systems data. We argue that research ethics committee (RECs) should play an important role in this developing field. Finally, we propose a framework for the regulation of digital participatory disease surveillance systems, which strengthens protection of participants' data and privacy, while promoting the concept of justice.

7.3 Consent

Traditional Informed Consent in Internet-Based Surveillance

In public health surveillance, there are 2 antithetical forces. Although these systems pursue the improvement of population health through surveillance of diseases (such as in the case of Influenzanet, providing protection for vulnerable populations at risk of serious adverse outcomes of influenza infection), they must also safeguard individuals against any abuse of their data by researchers [30,31]. To strike a balance between the pursuit of societal welfare and protection of individual rights, consent from participants plays a fundamental role. Originating from the necessity to protect research subjects both physically and mentally, written informed consent is traditionally obtained for medical research, and its importance has emerged even more in the current era of data protection [32]. However, this type of consent seems to be poorly adapted to the collection and use of digital data in public health surveillance [2].

In light of the inadequacy of traditional informed consent for participatory public health surveillance, 1 potential response is to reject the need for further consent in these types of studies [33] because of the fact that participants enroll on their own and not in continued medical contact. The "no problem" solution rests also on the assumption that consent is not necessary, as in public health surveillance, individual interests may be put aside to protect the public good [33]. For example, in the United States, the issue of consent in public health surveillance is circumvented by considering the latter as public health practice instead of

research, thus exempting it from institutional review boards' approval [34] and in most cases, of traditional informed consent requirements. Indeed, a participatory disease surveillance platform active in the United States, *Flu Near You*, received a waiver for informed consent [35]. However, this approach may not be the best solution for participatory disease surveillance where data are actively generated by participants, underscoring the urgent need to adapt the traditional model of informed consent more adequately to this type of surveillance system.

Informed consent in research was originally designed for studies involving a limited number of participants where it was practically and financially feasible for researchers to engage participants, provide details about the research, and obtain written informed consent before the beginning of the study [32]. A further problem with traditional informed consent is that it was designed to authorize the use of data only by those subjects and for those purposes according to which the data had been originally collected. It was, however, not intended to also cover retrospective research on samples or data. In the case of *big data* for surveillance of infectious diseases, which is often retrospective in nature (2 out of the 3 electronic data sources are medical encounter and nonhealth digital data) [12], obtaining traditional informed consent proves problematic, as it requires disclosing all potential risks of primary and retrospective research, but the latter are usually unknown at the time when data are collected [32].

There have been substantial efforts made by some Influenzanet national platforms at their outset to ensure some form of personal interaction with their participants to better explain the nature of the surveillance system. In 2003, the original Belgian/Dutch platform, called *de Grote Griepmeting* (ie, the Great Influenza Survey), received a lot of media attention, which led to the registration of tens of thousands of participants in 2003/2004. The participants' age distribution from youth to the elderly and their wide geographic spread and different levels of education made de Grote Griepmeting more accurate and quicker to signal the onset of a flu epidemic than the general practitioners' surveillance system organized by Netherlands Institute for Health Services Research (NIVEL). The Belgian/Dutch research team invited participants for an information, question and answer session, where they were provided with notes on the rationale of the survey questions to the flu survey study. Moreover, a forum was also created where participants could ask any remaining questions, and when specific virology questions were asked, consultation would follow with partners from the NIVEL and the National Institute for Health and Environment in the Netherlands. The team managed to answer all incoming questions from participants by email and during various local, national, and regional live radio interviews from people listening in. In 2009, the team started a public community on Facebook named *De Grote Griepmeting–Influenzanet*, where members would have their questions

answered by the team. Such measures (ie, the information session and the team answering all additional questions received via email, through their forum, on the social platform Facebook, and during radio interviews) could be viewed as an equivalent solution to obtaining the informed consent of these participants (De Grote Griepmeting, email communication, April 3, 2018 and February 7, 2019).

The inadequacies of traditional informed consent have led to the development of many other ethically acceptable solutions. For instance, in retrospective research where risks are minimal, consent would not be necessary as long as the right to opt out and the right to be forgotten are preserved and enforced [36,37]. Alternatively, the requirement of informed consent upholds but is paired with waivers, which dispense researchers from requesting consent for secondary use of data, if the recontacting and reconsenting are unfeasible or would lead to nonrepresentative samples [38].

Another alternative to traditional informed consent is an *extended* version of consent, which is more suitable to public health/*big data* research, known as broad or general consent [2,32,39-41]. The key difference between traditional and broad or general consent is that data subjects provide their consent for entire classes of research [42]. This extended form of consent differs from blanket consent as data subjects do not give permission for any use of their data but rather define in broad terms the purposes of use [42]. Moreover, broad consent is only considered acceptable if 2 criteria are met. First, every new study needs to be approved by an REC or another competent entity [43]. Second, the right of participants to withdraw their consent at any time has to be maintained [32,44]. Despite the presence of these safeguards, consensus on whether broad consent can be considered truly informed is lacking [45-47]. The informative nature of broad consent rests on the assumption that autonomy is protected, as REC approval is necessary, and strategies to regularly update the data donor on ongoing opt-out opportunities are devised [48]. Furthermore, any modification to the research should automatically lead to reconsenting procedures [49]. However, broad consent cannot be entirely informed because of the unspecified nature of future research [48]. Although broad consent seems suitable for secondary uses in public health research involving digital communities of volunteer citizens or *big data*, it is uncertain whether broad consent represents the best solution in terms of respect for autonomy. Given the issues raised by broad consent and the fact that it requires initial face-to-face contact, seeking consent electronically could be an ethically satisfying alternative to traditional informed consent.

Electronic Consent, An Adaptation of Traditional Informed Consent

Electronic consent (e-consent) implies that participants give informed consent using an information technology (eg, digital technologies). In this sense, e-consent does not represent a new form of consent but simply an adaptation of informed consent to the electronic environment [50]. E-consent is currently being used in the Influenzanet Consortium and in similar participatory surveillance platforms such as *Flu Tracking* (Australia and New Zealand) as a valid form of consent for participants. Data subjects agree to the conditions, terms, and privacy policies when registering on their respective national platforms [16,51].

Although e-consent offers the substantial benefit of a tailored fit to the digital environment, it also has some inherent problems. A unique feature of internet-based research is the absence of personal interaction between researchers and participants, where researchers would traditionally be able to provide individually tailored information and answer any question participants might ask concerning the study and the collection of health data. Therefore, one of the major risks posed by e-consent is the provision of consent through automatic processes in the digital world, as parties are not directly involved. The provision of consent is rather based on a set of computer rules determining whether access to an individual's data by researchers could be granted on reasonable grounds [50]. For this reason, it is possible that participants provide their consent without fully understanding—or even reading—the information, terms, and conditions that data collectors provide by simply clicking the relevant buttons in the digital forms [52]. We thus recommend that several precautions ought to be implemented when e-consent is obtained. For instance, e-consent should be designed in such a way that information is delivered through a simple PDF file where participants digitally put their signature (instead of clicking a button) to increase the likelihood of the document being read. Alternatively, other possibilities offered by information technology could be exploited to help verify participants' understanding of the information provided during the e-consent process. These include tools such as the use of audio files, PowerPoint presentations, videos, pictures, or gamification (for instance, through quizzes and animation) [53].

Though the above recommendations could foster the informed nature of e-consent, the lack of personal interaction between study participants and researchers remains. Therefore, a properly implemented e-consent would be particularly beneficial in those studies where it is impossible to provide individual counseling and where the conditions, terms, and privacy policies would otherwise not be read [54,55]. In addition, one might even argue that participants are potentially less likely to consent under undue influence or constraints because of limited interaction with researchers. They can thus easily decline consent by signing out from the digital platforms

whenever they feel the need to do so [52]. In this regard, e-consent increases the autonomy of participants.

Nonetheless, a further challenge raised by the absence of face-to-face contact is how to ensure that participants have the required legal capacity to legitimately give their e-consent after they electronically authenticate [52], because of the difficulty of verifying the participants' identity. In this sense, even if measures were taken—such as quizzes or questionnaires—to ensure that participants have understood the research information, there would be little guarantee that those quizzes or study questionnaires are actually being completed by data subjects. A potential solution to this authentication issue could come from advances made in biometric identification technologies, commonly used for security purposes [52]. For instance, the use of face recognition technology [56] on computers and mobile phones or fingerprint recognition sensor in smartphones (eg, Touch ID by Apple Inc) could be used to verify the identity of participants throughout the process of e-consent. However, processing biometric data raises additional ethical and legal issues, in particular with respect to privacy. Biometric data, similar to genetic material, carries biological traits that are unique to data subjects and which could be easily used to reidentify them [57]. However, it must be noted that, although entailing sensitive personal information, the processing of biometric data can be lawful even without subjects' consent if processing serves the public interest or scientific or statistical research purposes (eg, Article 9.2.(j) of the European General Data Protection Regulation, GDPR) [58]. Ethical concerns with respect to biometric data might also be mitigated if, for example, the gathered biometric information was stored locally on the participants' computers or mobile devices and not transmitted to the research team or any third party.

In the case of multinational studies such as Influenzanet, an adequately implemented e-consent could consist of a standard informed consent information form [16], as a single PDF file, being delivered to participants at the time of their registration on the digital platform and followed by a new document each and every time new information is added on the single country websites. The information provided would have to be reader-friendly and succinctly summarized, thus nudging participants to read it thoroughly. Awareness of all potential ramifications because of their participation could be further improved through the provision of quiz questions. Grading of these quiz questions could then serve as a proxy to ensure adequate understanding of the informed consent information. This method has been employed at the *Harvard Personal Genome Project* (PGP) [59]. Participants were even provided with a study guide and were required to pass an enrollment test to be considered for the project. This additional burden to participation, which is justified for genetic research (*genetic exceptionalism*), should

nonetheless remain minimal for Influenzanet to retain engagement of its participants. This is supported by the mildly sensitive nature of the gathered information and the low risks associated with this kind of surveillance. Indeed, the enrollment examination for the PGP was the main barrier to participation, with almost 60% of its users dropping out [59]. Digital signature of the consent form could also be a more personalized alternative, and it would also provide additional evidence on the identity of the participant, which altogether would enhance the informed nature of this e-consent procedure.

7.4 Protection of Subjects' Privacy

Epidemics forecasting studies and other public health research often gather useful and sensitive data on their participants, potentially interfering with their privacy. In the case of Influenzanet, protection of participants' privacy is secured by data anonymization and the use of a centralized database [18].

One might argue that full anonymization is not necessary for some public health surveillance, as part of the collected data is only mildly sensitive (eg, age group and gender) and thus poses only a minor threat to the fundamental rights and the privacy of participants even in case of misuse [60]. However, even nonpersonal information could be used to reveal much more sensitive information on data subjects if the former is coupled with additional geographical information, which is often collected by public health surveillance systems [61]. For instance, 1 of the core functions of Influenzanet is to map cases of *ILI* for the identification of hotspots of influenza outbreaks to model disease progression and implement effective prevention strategies. This spatiotemporal dimension of collected health data can enhance the privacy-invasive nature of epidemics forecasting research such as Influenzanet [62]. The collection of sensor and usage data from smartphones adds additional behavioral and context information, which, as shown in related work [15], has the potential to improve forecasting and risk analysis. Despite these potential benefits, even apparently nonpersonal data, such as a list of installed apps, can be an additional risk to the participant's privacy [63]. The Consortium took great care in protecting the privacy of its participants. In case of sensor data, information is processed directly on the user's device and only transmitted to Influenzanet in anonymized and highly aggregated form [15]. Location information of reported cases is, for example, never mapped to the individual level but rather to the postal code level [17], with only the aggregate number of cases shown. Some platforms went even further by randomizing virtual locations around the center of a large number of postal code areas taken together (eg, the De Grote Griepmeting platform). The grouping of postal codes areas was paramount for better protection of the privacy

of participants, for example, in the case of a single participant in a postal code area (De Grote Griepmeting, email communication, April 3, 2018).

However, with increasing technological capabilities to integrate and analyze health data with local data, there are risks of leakage of sensitive information concerning participants' locations, which may lead to stigmatization of the particular locations as well as residents [62,64]. Even with full anonymization, cross-referencing of essential data gathered for epidemics research purposes (eg, sex, age, and medical conditions) with other databases could eventually lead to reidentification of data subjects [2,65]. For instance, 2 researchers showed it was feasible to reidentify individuals by matching a deidentified database on *Netflix* movie recommendations to available Web-based information (eg, Internet Movie Database) [66]. Hence, anonymization per se is not a sufficient measure to adequately protect privacy. The long life span of some anonymized datasets, which is often the case with epidemics forecasting studies, de facto increases the risks of reidentification and privacy breaches through repetitive data enrichment over time [54]. Consequently, reidentification should be considered a real risk for data subjects [2] even in case of anonymization. It is, thus, paramount to ensure ethical and accountable sharing of anonymized datasets between research institutions and to combine anonymization with other adequate data security measures to prevent misuse of data and unauthorized reidentification.

7.5 Justice

The concept of justice in research ethics is fully embodied by the policies of the Influenzanet Consortium as participation is free, open, voluntary, and nondiscriminative of any resident of the respective countries (except for Sweden, where some representativeness and comparison purposes are guaranteed by allowing participation through invitations only) [17]. This ensures a fair distribution of risks and benefits to all research participants and the public at large. If epidemics forecasting studies keep up with the high standards in terms of justice (ie, participation to the surveillance system is free, open, voluntary, and nondiscriminative) followed by the Influenzanet Consortium and similar platforms such as Flu Near You [35], the only remaining challenge would be dealing with those limitations on participation that are inherent to digital technologies. These limitations are commonly referred in the literature as the "*digital divide*" [2,67], and they concern both access to and proficiency with digital technologies. Access to digital technologies is also a product of many sociodemographic elements such as age, educational level of participants, ethnic groups, and their socioeconomic status [68,69]. This is reflected in the data collected by Influenzanet, which present an

underrepresentation of younger age and elderly groups, an overrepresentation of the middle age group for both genders, and a higher educational level of participants in comparison with the general population [70]. In this respect, it could be claimed that epidemics forecasting studies such as Influenzanet are potentially empowering a more dynamic, informed societal group with a penchant for digital technology, whereas at the same time perpetuating the health inequalities between others [2]. However, it must also be stressed that public health surveillance benefits the public at large and not exclusively the participants. Public health surveillance, like biomedical research, is a public good, as the health benefits resulting from its interventions (based on knowledge generated from data subjects) are ultimately going to be shared with society [71]. In addition, the digital divide is decreasing annually, with technology becoming more and more pervasive [72]. Nevertheless, concerns about justice can be avoided only if results and disease prevention strategies are shared evenly and on a regular basis among all societal groups, something which the Influenzanet Consortium is promoting (eg, weekly national surveillance bulletins, regular press releases during the study, and radio broadcasting) [73,74]. Such regular results dissemination initiatives undeniably help in ensuring that expected benefits of research are shared more equally between societal groups. This could be further improved by granting access to more targeted and granular information on influenza activity to nonparticipants [74] under the concept of solidarity [75]. Such measures would allow a better protection of society as the spread of an influenza epidemic is an individual as well as a collective concern.

7.6 Capacity Building for Research Ethics Committees

Influenzanet and similar systems are faced with multifocal ethical and legal issues. For the safeguard of data subjects, appropriate oversight and specific regulation might be needed in the future. Currently, such oversight is beyond the governance capacity of RECs, as technological advances outpace national regulatory frameworks and undermine the definitions of those concepts—such as “anonymization,” “encryption,” and “personally identifiable information” [55]—upon which RECs rely. However, RECs should be actively involved in the design and implementation of public health research involving digital communities of volunteer citizens or big data. These RECs need to act as safety nets to fill the gaps of the current regulatory framework, which often dates back to an era where modern computational and technological capabilities were not foreseeable [55]. In this perspective, we recommend RECs to undergo interdisciplinary capacity building in those innovative research methods through mutual exchange of information and training with citizen science experts, big data researchers, data scientists, ethicists, legal experts, and sociologists. This would allow the identification of ethical

and legal grey zones. Stakeholders could further anticipate potential conflicting situations resulting from the enactment of new legislation. This appears even more urgent as we have entered the GDPR era. This regulation came into force in May 2018, to replace the EU Data Protection Directive 95/46/EC [1,76]. The GDPR tries to harmonize EU data protection laws with the goal of guaranteeing the same level of freedom and protection to EU citizens, while protecting personal data during cross-border sharing with international organizations and third countries [76]. This legislative reform is likely to have a considerable impact on consent requirements and exemptions from obtaining consent [1]. This could affect the expected benefits that *big data* can bring to society by increasing the regulatory burden on public health surveillance studies [1]. Furthermore, as stressed in the study by Mittelstadt, the GDPR classifies "data concerning health" as a "special category of personal data" [77]. As this category includes any personal data that reveals information on the health status (physical or mental) of participants [77], health-related information gathered from Influenzanet participants or similar epidemics forecasting studies might—until properly anonymized—fall in this special category. It is, thus, possible that detailed limitations to health data usage are imposed in the future because of the protective stance endorsed by the GDPR [77]. Therefore, interdisciplinary capacity building and acquaintance of RECs with this innovative and developing research field will be paramount to proactively ensure an adequate protection of data subjects while preventing the development of additional research barriers. Such barriers could undermine the excellent contribution to the preservation of public health made by epidemics forecasting systems such as Influenzanet.

7.7 Ethical Framework for the Regulation of Participatory Disease Surveillance Systems

We propose the following 4 components ethical framework to provide guidance on how to ensure an adequate ethical oversight of participatory disease surveillance systems while safeguarding participants' privacy and eliminating barriers to the work of these surveillance platforms (Table 1).

Table 1 Ethical framework for the regulation of participatory disease surveillance systems

Principle	Ethical Component	Considerations
Autonomy of participants	Electronic consent	Standard, reader-friendly, and multilingual informed consent form with succinctly summarized information (eg, as a single PDF file) delivered at the time of registration and each time new information is added to the digital platforms
		Informed nature of consent can be fostered through the provision of a few quiz questions to reduce the risk of participants simply "clicking through" the consent process
		Require digital signature of the consent form to incentivize participants to read the information form and as evidence of their identity
		Making participants aware of the fact that despite best effort to protect their privacy, the residual risk of a privacy leak cannot be ruled out
Nonmaleficence	Protection of participants' privacy	Anonymization of participants' data should be combined with other data security measures such as a highly protected centralized database for storage of participants' data
		Location data of participants should never be mapped to the individual level but rather to the postal code level to reduce the risk of reidentification in case of rare value entries
		Sensor data from mobile phones should only be transmitted in anonymized and highly aggregated form
		Ensure ethical and accountable sharing of anonymized datasets between research institutions to reduce reidentification risks for participants through database triangulation
Justice	Access to information on disease activity and prevention strategies	Free, open, and nondiscriminative participation should be offered to members of the general public
		Disease prevention strategies and results obtained through the participatory surveillance platforms should be disseminated on a regular basis to members of the public through various means
Beneficence and nonmaleficence	Research ethics committees (RECs)	Interdisciplinary capacity building of RECs is required to keep up with technological advances, thereby ensuring an adequate protection of data subjects
		RECs should play a proactive role in the design and implementation of public health research involving digital communities of volunteer citizens
		RECs should act as safety nets to prevent barriers to public health surveillance by identifying ethico-legal grey zones and anticipate potential conflicting situations resulting from the evolving legal landscape

7.8 Conclusions

In the developing field of participatory disease surveillance systems, the main ethical dilemma is how to ensure adequate protection of data subjects while at the same time obtaining the full benefits that public health surveillance directly involving digital communities of citizens could bring. In this complex situation, 1 of the key ethical safeguards proposed in our framework is a properly implemented e-consent. To pursue this objective, national platforms of the Influenzanet Consortium will put continuous effort in enhancing and adequately developing their e-consent procedures. Current e-consent procedures could be improved by providing standard, reader-friendly, multilingual information about the study, participants' rights, the risks associated with their participation, and, in addition, a short series of quiz questions to verify proper understanding of the potential benefits and risks. Furthermore, requiring participants to digitally sign the Web-based consent form could both serve as a motivation for them to read properly the information provided and as a solution to allow personal identification. However, such additional burdens of participation need to remain minimal to ensure the sustainability of the platforms.

Acknowledgments

LDG, AM, TW, and BSE acknowledge the financial support provided by the Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356). The views expressed in this paper are those of the authors and not necessarily those of the funder. The authors would like to thank Ricardo Mexia for his contribution to the conceptualization of the manuscript. LDG thanks Christopher Poppe and Maddalena Favaretto for their help. OWM was affiliated with ETH Zurich, the Swiss Federal Institute of Technology when the work for the publication was carried out and is currently affiliated with Novartis Pharma AG.

Authors' Contributions

All authors (except PL) contributed to the conceptualization of this paper. All authors contributed to the writing, editing, and critical evaluation of the manuscript. They approved the submission of the final version of the manuscript.

Conflicts of Interest

DP, CK, CK, CG, MH, OWM, PL, and AF are members of the Influenzanet Consortium. MH is cofounder of coneno, a software development company which is working together with the Influenzanet consortium since December 2018 as a technology partner to develop a new open source Influenzanet platform to be launched in the future. The development of the new platform is, at its current state mostly driven by volunteer work and partly funded by ISI foundation (part of Influenzanet Consortium).

Abbreviations

e-consent: electronic consent

EISN: European Influenza Surveillance Network

EU: European Union

GDPR: General Data Protection Regulation

GFT: Google Flu Trends

ILI: influenza-like illness

NIVEL: Netherlands Institute for Health Services Research

PGP: Personal Genome Project

REC: research ethics committee

7.9 References

1. Mostert M, Bredenoord AL, Biesaart MC, van Delden JJ. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet 2016 Jul;24(7):956-960 [doi:10.1038/ejhg.2015.239] [Medline: 26554881]
2. Thiel MJ, Parkin J, Dratwa J, Halila R, Palazzani L, Górski A. Publications Office of the European Union. Luxembourg; 2016 Mar 18. The ethical implications of new health technologies and citizen participation URL:https://publications.europa.eu/en/publication-detail/-/publication/e86c21fa-ef2f-11e5-852901aa75ed71a1 [accessed 2017-10-01] [WebCite Cache ID 72Zh0cB7k]
3. International Telecommunication Union. 2018 Dec 07. ITU releases 2018 global and regional ICT estimates URL:https://www.itu.int/en/mediacentre/Pages/2018-PR40.aspx [accessed 2019-04-18] [WebCite Cache ID 77iofmwuq]
4. Aitken M, Clancy B, Nass D. IQVIA. 2017 Nov 07. The Growing Value of Digital Health URL:https://www.iqvia.com/ institute/reports/the-growing-value-of-digital-health [accessed 2019-04-18] [WebCite Cache ID 77iphhud5]
5. Klingler C, Silva DS, Schuermann C, Reis AA, Saxena A, Strech D. Ethical issues in public health surveillance: a systematic qualitative review. BMC Public Health 2017 Apr 4;17:295 [doi: 10.1186/s12889-017-4200-4] [Medline: 28376752]
6. Chan EH, Brewer TF, Madoff LC, Pollack MP, Sonricker AL, Keller M, et al. Global capacity for emerging infectious disease detection. Proc Natl Acad Sci U S A 2010 Dec 14;107(50):21701-21706 [doi:10.1073/pnas.1006219107] [Medline: 21115835]
7. Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. N Engl J Med 2013 Aug 1;369(5):401-404. [doi: 10.1056/NEJMp1307752] [Medline: 23822655]

8. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
9. Centers for Disease Control and Prevention. 2010 Jul 29. Distinguishing public health research and public health nonresearch. URL:https://www.cdc.gov/od/science/integrity/docs/cdc-policy-distinguishing-public-health-research-nonresearch.pdf [accessed 2019-04-18] [WebCite Cache ID 77itmHtGr]
10. Smolinski MS, Crawley AW, Olsen JM, Jayaraman T, Libel M. Participatory disease surveillance: engaging communities directly in reporting, monitoring, and responding to health threats. JMIR Public Health Surveill 2017 Oct 11;3(4):e62 [doi: 10.2196/publichealth.7540] [Medline: 29021131]
11. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012 Jul 26;8(7):e1002616 [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]
12. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. J Infect Dis 2016 Dec 1;214(suppl_4):S375-S379 [doi: 10.1093/infdis/jiw400] [Medline: 28830113]
13. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
14. European Commission. 2014 Nov 27. EPIWORK: Ebola forecasting uses model developed by EU project URL:https://ec.europa.eu/programmes/horizon2020/en/news/epiwork-ebola-forecasting-uses-model-developed-eu-project [accessed 2018-06-01] [WebCite Cache ID 72Zjbbbft]
15. Hirsch M, Woolley-Meza O, Paolotti D, Flahault A, Lukowicz P. grippeNET App: Enhancing Participatory Influenza Monitoring Through Mobile Phone Sensors. : ACM; 2018 Oct 8 Presented at: The ACM International Joint Conference and International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers; October 08-12, 2018; Singapore p. 833-841. [doi: 10.1145/3267305.3274171]
16. Geneviève LD, Wangmo T, Dietrich D, Woolley-Meza O, Flahault A, Elger BS. Research ethics in the European Influenzanet Consortium: scoping review. JMIR Public Health Surveill 2018 Oct 10;4(4):e67. [Medline: 30305258]
17. Koppeschaar CE, Colizza V, Guerrisi C, Turbelin C, Duggan J, Edmunds WJ, et al. Influenzanet: citizens among 10 countries collaborating to monitor influenza in Europe.

JMIR Public Health Surveill 2017 Sep 19;3(3):e66 [doi: 10.2196/publichealth.7429] [Medline: 28928112]

18. Paolotti D, Carnahan A, Colizza V, Eames K, Edmunds J, Gomes G, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. Clin Microbiol Infect 2014 Jan;20(1):17-21 [doi: 10.1111/1469-0691.12477] [Medline: 24350723]
19. Simonsen L, Gog JR, Olson D, Viboud C. Infectious disease surveillance in the big data era: towards faster and locally relevant systems. J Infect Dis 2016;214(suppl_4):S380-S385. [Medline: 28830112]
20. van Noort SP, Codeço CT, Koppeschaar CE, van Ranst M, Paolotti D, Gomes MG. Ten-year performance of Influenzanet: ILI time series, risks, vaccine effects, and care-seeking behaviour. Epidemics 2015 Dec;13:28-36 [doi:10.1016/j.epidem.2015.05.001] [Medline: 26616039]
21. Friesema IH, Koppeschaar CE, Donker GA, Dijkstra F, van Noort SP, Smallenburg R, et al. Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands. Vaccine 2009 Oct 23;27(45):6353-6357. [doi: 10.1016/j.vaccine.2009.05.042] [Medline: 19840672]
22. Debin M, Colizza V, Blanchon T, Hanslik T, Turbelin C, Falchi A. Effectiveness of 2012-2013 influenza vaccine against influenza-like illness in general population: estimation in a French web-based cohort. Hum Vaccin Immunother 2014 Mar 1;10(3):536-543 [Medline: 24343049]
23. Boiron K, Sarazin M, Debin M, Raude J, Rossignol L, Guerrisi C, et al. Opinion about seasonal influenza vaccination among the general population 3 years after the A(H1N1)pdm2009 influenza pandemic. Vaccine 2015 Nov 27;33(48):6849-6854. [Medline: 26322844]
24. Debin M, Turbelin C, Blanchon T, Bonmarin I, Falchi A, Hanslik T, et al. Evaluating the feasibility and participants' representativeness of an online nationwide surveillance system for influenza in France. PLoS One 2013 Sep 11;8(9):e73675 [doi: 10.1371/journal.pone.0073675] [Medline: 24040020]
25. Vayena E, Mastroianni A, Kahn J. Ethical issues in health research with novel online sources. Am J Public Health 2012 Dec;102(12):2225-2230. [doi: 10.2105/AJPH.2012.300813] [Medline: 23078484]

26. Myers J, Frieden T, Bherwani K, Henning K. Ethics in public health research: privacy and public health at risk: public health confidentiality in the digital age. Am J Public Health 2008 May;98(5):793-801. [doi: 10.2105/AJPH.2006.107706] [Medline: 18382010]
27. World Health Organization. WHO Guidelines on Ethical Issues in Public Health Surveillance. Geneva: World Health Organization; 2017.
28. Salerno J, Knoppers BM, Lee LM, Hlaing WM, Goodman KW. Ethics, big data and computing in epidemiology and public health. Ann Epidemiol 2017 May;27(5):297-301. [doi: 10.1016/j.annepidem.2017.05.002] [Medline: 28595734]
29. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb 9;11(2):e1003904 [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
30. Woolley JP. How data are transforming the landscape of biomedical ethics: the need for ELSI metadata on consent. In: Mittelstadt B, Floridi L, editors. The Ethics of Biomedical Big Data. Cham: Springer; 2016:171-197.
31. Quinn SC. Ethics in public health research: protecting human subjects: the role of community advisory boards. Am J Public Health 2004 Jun;94(6):918-922. [Medline: 15249289]
32. Elger BS, Caplan AL. Consent and anonymization in research involving biobanks: differing terms and norms present serious barriers to an international framework. EMBO Rep 2006 Jul;7(7):661-666 [doi: 10.1038/sj.embor.7400740] [Medline: 16819458]
33. Ploug T, Holm S. Meta consent–a flexible solution to the problem of secondary use of health data. Bioethics 2016;30(9):721-732 [doi: 10.1111/bioe.12286] [Medline: 27628305]
34. Otto JL, Holodniy M, DeFraites RF. Public health practice is not research. Am J Public Health 2014 Apr;104(4):596-602. [doi: 10.2105/AJPH.2013.301663] [Medline: 24524499]
35. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015 Oct;105(10):2124-2130. [doi: 10.2105/AJPH.2015.302696] [Medline: 26270299]
36. Olsen J. Meta consent - a workable procedure in the area of Big Data? Br Med J 2015 Jun 1;350:h2146 [FREE Full text] [doi: 10.1136/bmj.h2146]

37. Elger B, Mauron A. A presumed-consent model for regulating informed consent of genetic research involving DNA banking. In: Knoppers BM, editor. Populations and Genetics: Legal Socio-Ethical Perspectives. Leiden: Martinus Nijhoff; 2003:269-295.
38. Gefenas E, Dranseika V, Cekanauskaite A, Serepkaite J. Research on human biological materials: what consent is needed, and when. In: Lenk C, Sándor J, Gordijn B, editors. Biobanks and Tissue Research: The Public, the Patient and the Regulation. Dordrecht: Springer; 2011:95-110.
39. Ioannidis JP. Informed consent, big data, and the oxymoron of research that is not research. Am J Bioeth 2013;13(4):40-42. [doi: 10.1080/15265161.2013.768864] [Medline: 23514395]
40. Colledge F, Elger BS. Impossible, impractical, and non-identifiable? New criteria regarding consent for human tissue research in the Declaration of Helsinki. Biopreserv Biobank 2013 Jun;11(3):149-152. [doi: 10.1089/bio.2013.0006] [Medline: 24850091]
41. Colledge F, Persson K, Elger B, Shaw D. Sample and data sharing barriers in biobanking: consent, committees, and compromises. Ann Diagn Pathol 2014 Apr;18(2):78-81. [doi: 10.1016/j.anndiagpath.2013.12.002] [Medline: 24485935]
42. Helgesson G. In defense of broad consent. Camb Q Healthc Ethics 2012 Jan;21(1):40-50. [doi: 10.1017/S096318011100048X] [Medline: 22152457]
43. Hansson MG, Dillner J, Bartram CR, Carlson JA, Helgesson G. Should donors be allowed to give broad consent to future biobank research? Lancet Oncol 2006 Mar;7(3):266-269. [doi: 10.1016/S1470-2045(06)70618-0] [Medline: 16510336]
44. Elger B. Withdrawal of consent and destruction of samples. In: Elger B, Biller-Andorno A, Mauron A, Capron AM, editors. Ethical Isues in Governing Biobanks: Global Perspectives. London: Ashgate; 2008:131-166.
45. Sheehan M. Can broad consent be informed consent? Public Health Ethics 2011 Nov;4(3):226-235. [doi: 10.1093/phe/phr020] [Medline: 22102849]
46. Elger B. Consent and use of samples. In: Elger B, Biller-Andorno N, Mauron A, Capron AM, editors. Ethical Issues in Governing Biobanks: Global Perspectives. London: Ashgate; 2008:57-88.
47. Elger B. Consent to research involving human biological samples obtained during medical care. In: Elger B, Biller-Andorno N, Mauron A, Capron AM, editors. Ethical Issues in Governing Biobanks: Global Perspectives. London: Ashgate; 2008:89-120.

48. Steinsbekk KS, Myskja BK, Solberg B. Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? Eur J Hum Genet 2013 Sep;21(9):897-902 [doi: 10.1038/ejhg.2012.282] [Medline: 23299918]
49. Steinsbekk KS, Solberg B. Biobanks - when is re-consent necessary? Public Health Ethics 2011 Nov 2;4(3):236-250. [doi: 10.1093/phe/phr031]
50. Coiera E, Clarke R. e-Consent: the design and implementation of consumer consent mechanisms in an electronic environment. J Am Med Inform Assoc 2004;11(2):129-140. [Medline: 14662803]
51. FluTracking. About Flutracking URL:https://www.flutracking.net/Join [accessed 2019-01-24] [WebCite Cache ID 75f0e69fe]
52. Grady C, Cummings SR, Rowbotham MC, McConnell MV, Ashley EA, Kang G. Informed consent. N Engl J Med 2017;376(9):856-867. [doi: 10.1056/NEJMra1603773] [Medline: 28249147]
53. Simon CM, Klein DW, Schartz HA. Traditional and electronic informed consent for biobanking: a survey of U.S. biobanks. Biopreserv Biobank 2014 Dec;12(6):423-429. [doi: 10.1089/bio.2014.0045] [Medline: 25496155]
54. Metcalf J, Crawford K. Where are human subjects in big data research? The emerging ethics divide. Big Data Soc 2016 Jun 17;3(1):1-14. [doi: 10.1177/2053951716650211]
55. Tene O, Polonetsky J. Big data for all: privacy and user control in the age of analytics. Nw J Tech Intell Prop 2013;11(5):240-273
56. Voth D. Face recognition technology. IEEE Intell Syst 2003 May;18(3):4-7. [doi: 10.1109/MIS.2003.1200719] [Medline: 27295638]
57. Alterman A. "A piece of yourself": ethical issues in biometric identification. Ethics Inf Technol 2003;5(3):139-150 [doi: 10.1023/B:ETIN.0000006918.22060.1f]
58. EUR-Lex. 2016 Apr 27. Regulation (EU) 2016/679 of the European Parliament and of the Council.URL:https://eurlex.europa.eu/legalcontent/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC [accessed 2018-06-03] [WebCite Cache ID 72ZlDVdRz]
59. Ball MP, Bobe JR, Chou MF, Clegg T, Estep PW, Lunshof JE, et al. Harvard Personal Genome Project: lessons from participatory public research. Genome Med 2014 Feb 28;6(2):10 [doi: 10.1186/gm527] [Medline: 24713084]
60. Article 29 Data Protection Working Party. 2011 Apr 4. Advice paper on special categories of data ("sensitive data").URL:https://ec.europa.eu/justice/article-

29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf [accessed 2019-04-19] [WebCite Cache ID 77kFCvEZg]

61. Zwitter A. Big data ethics. Big Data Soc 2014 Nov 20;1(2):1-6. [doi: 10.1177/2053951714559253]
62. Ölvingson C, Hallberg J, Timpka T, Lindqvist K. Ethical issues in public health informatics: implications for system design when sharing geographic information. J Biomed Inform 2002 Jun;35(3):178-185 [Medline: 12669981]
63. Welke P, Andone I, Blaszkiewicz K, Markowetz A. Differentiating Smartphone Users by App Usage. 2016 Presented at: UBICOMP'16; September 12-16, 2016; Heidelberg, Germany p. 519-523. [doi: 10.1145/2971648.2971707]
64. Vicente CR, Freni D, Bettini C, Jensen CS. Location-related privacy in geo-social networks. IEEE Internet Comput 2011 May;15(3):20-27. [doi: 10.1109/MIC.2011.29]
65. Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for responsible sharing of clinical trial data. N Engl J Med 2013 Oct 24;369(17):1651-1658. [Medline: 24144394]
66. Narayanan A, Shmatikov V. Robust De-anonymization of Large Sparse Datasets. 2008 Presented at: IEEE Symposium on Security and Privacy; May 18-22, 2008; Oakland, CA, USA p. 111-125. [doi: 10.1109/SP.2008.33]
67. van Dijk J, Hacker K. The digital divide as a complex and dynamic phenomenon. Informa Soc 2003;19(4):315-326. [doi:10.1080/01972240309487]
68. Friemel TN. The digital divide has grown old: determinants of a digital divide among seniors. New Media Soc 2014 Jun 12;18(2):313-331. [doi: 10.1177/1461444814538648]
69. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [doi: 10.2196/jmir.3117] [Medline: 25048379]
70. Cantarelli P, Debin M, Turbelin C, Poletto C, Blanchon T, Falchi A, et al. The representativeness of a European multi-center network for influenza-like-illness participatory surveillance. BMC Public Health 2014 Sep 20;14:984 [doi: 10.1186/1471-2458-14-984] [Medline: 25240865]
71. Schaefer GO, Emanuel EJ, Wertheimer A. The obligation to participate in biomedical research. J Am Med Assoc 2009 Jul 1;302(1):67-72. [Medline: 19567441]

72. International Telecommunication Union. Geneva, Switzerland: ITU; 2017 Jul. ICT FACTS AND FIGURES 2017 URL:https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf [accessed 2019-04-19] [WebCite Cache ID 77kIi654d]
73. Perrotta D, Bella A, Rizzo C, Paolotti D. Participatory online surveillance as a supplementary tool to sentinel doctors for influenza-like illness surveillance in Italy. PLoS One 2017 Jan;12(1):e0169801 [doi:10.1371/journal.pone.0169801] [Medline: 28076411]
74. Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, et al. Participatory syndromic surveillance of influenza in Europe. J Infect Dis 2016 Dec 1;214(suppl_4):S386-S392. [doi: 10.1093/infdis/jiw280] [Medline: 28830105]
75. Dawson A, Jennings B. The place of solidarity in public health ethics. Public Health Rev 2012 Jun 12;34(1):65-79. [doi:10.1007/BF03391656]
76. Preite F, Salardi S, Gesuita R, Villani S, Trerotoli P, Guardabasso V, et al. The new European regulation on personal data protection: significant aspects for data processing for scientific research purposes. Epidemiol Biostat Public Health 2017;14(2):e12286-1-e1228613. [doi: 10.2427/12286]
77. Mittelstadt B. Designing the health-related internet of things: ethical principles and guidelines. Information 2017;8(3):77.[doi: 10.3390/info8030077]

Chapter 8: Precision Public Health and Structural Racism in the United States: Promoting Health Equity in the COVID-19 Pandemic Response

Lester Darryl Geneviève[1], Andrea Martani[1], Tenzin Wangmo[1], Bernice Simone Elger[1,2]

[1]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland.

[2]University Center of Legal Medicine, University of Geneva, Geneva, Switzerland.

Citation: Geneviève LD, Martani A, Wangmo T, Elger BS. Precision Public Health and Structural Racism in the United States: Promoting Health Equity in the COVID-19 Pandemic Response. JMIR Public Health Surveill. 2022;8(3):e33277. DOI: 10.2196/33277 [CC BY 4.0]

"In order to get beyond racism, we must first take account of race. There is no other way. And in order to treat some persons equally, we must treat them differently." - Harry A. Blackmun

8.1 Abstract

The COVID-19 pandemic has revealed deeply entrenched structural inequalities that resulted in an excess of mortality and morbidity in certain racial and ethnic groups in the United States. Therefore, this paper examines from the US perspective how structural racism and defective data collection on racial and ethnic minorities can negatively influence the development of precision public health (PPH) approaches to tackle the ongoing COVID-19 pandemic. Importantly, the effects of structural and data racism on the development of fair and inclusive data-driven components of PPH interventions are discussed, such as with the use of machine learning algorithms to predict public health risks. The objective of this viewpoint is thus to inform public health policymaking with regard to the development of ethically sound PPH interventions against COVID-19. Particular attention is given to components of structural racism (eg, hospital segregation, implicit and organizational bias, digital divide, and sociopolitical influences) that are likely to hinder such approaches from achieving their social justice and health equity goals.

8.2 Structural Racism and Defective Data: Health Equity Threats for the Public Health Response Against COVID-19

Structural racism refers to "[a] system in which public policies, institutional practices, cultural representations, and other norms work in various, often reinforcing ways to perpetuate racial group inequity" [1]. Structural racism affects all social determinants of health. Its impact was accentuated and made more visible with the COVID-19 pandemic. Indeed, the US Centers for Disease Control and Prevention recognized the centrality of the social determinants of health in the disproportionate impact of the pandemic on racial and ethnic minority groups. In addition, they acknowledged the pervasive contributing influences of (structural) racism across each one of these determinants [2]. Undeniably, since the start of the pandemic, a cornucopia of examples suggesting the impact of structural racism has surfaced [3]. This was particularly evident in countries from which statistics are presently available, such as the United States, showing devastating consequences for the health of racial and ethnic minorities [4]. For instance, as of November 2021 in the United States, the risk of dying from COVID-19 for non-White groups (except for the Asian non-Hispanic community) was about twice that of White people [5]. Such inequities can be partly explained by structural vulnerabilities influencing access to and quality of care offered to these racial/ethnic minority groups.

The structural vulnerabilities of some racial and ethnic minority groups to COVID-19 stem from political and social influences, which were observed, for instance, to have a greater impact on their health than individual choices made [6]. These influences led to the normalization of

discrimination, stereotyping, and prejudices, and ultimately impacted racial minorities' health and access to quality care [7]. Indeed, studies have demonstrated the existence of implicit racial/ethnic biases among health care professionals, and their presumed impact on the quality of care due to suboptimal patient-provider interactions with racial/ethnic minorities (see review [8]). For instance, emerging evidence based on billing data started indicating disparities in COVID-19 testing among racial and ethnic groups, with African Americans being less likely to be offered COVID-19 testing than Whites, even when presenting with similar symptoms such as fever and cough [9]. Aside from individual health care professionals, health care institutions and organizations can also harbor implicit biases regarding their approach toward minority groups. It was highlighted that testing centers for COVID-19 in the United States were predominantly located in wealthy and White neighborhoods, which further limits access to health care for people located in poor neighborhoods [10]. These implicit provider and institutional/organizational biases could be an additional source of structural health disparities for ethnic and racial groups during the COVID-19 pandemic [11]. Indeed, average threshold metrics for public health interventions (eg, those used as indicators for closing or opening schools/businesses) are biased if COVID-19 testing is carried out mostly in wealthy neighborhoods—where viral transmission rates are actually lower—and deprived neighborhoods (where residents are predominantly minorities) are insufficiently tested [12].

Additionally, public health recommendations for containing the spread of COVID-19 have been particularly difficult for racial and ethnic minority groups to implement, partly due to the downstream influences of structural racism [3,13]. As rightly noted by Krieger [14], racial and ethnic minority groups constitute a good percentage of low-wage workers, making them more vulnerable to the effects of the pandemic. They often live in crowded multigenerational houses (often a consequence of racial residential segregation and redlining policies—two historical examples of structural racism [15,16]), with neither the possibility of working remotely from the safety of their homes nor adequate access to COVID-19 testing and treatment (eg, because minorities are less likely to be insured) [3,17].

Other indicators of how structural elements determine health inequity for racial/ethnic minorities during the COVID-19 pandemic have been highlighted by some seroprevalence studies. For instance, a large-scale nationwide study carried out in the United States showed higher seropositivity rates (2- to 3-fold higher) for SARS-CoV-2 antibodies in dialysis patients residing in Hispanic and non-Hispanic Black neighborhoods in comparison to those residing in White (non-Hispanic) neighborhoods, and a 2-fold higher seropositivity for those living in

poorer neighborhoods [18]. This study and others (eg, [19]) challenge the underlying assumption made elsewhere that “the risk of infection is homogeneous within the population” [19]. Such implicit assumptions are not only detrimental to the efforts made to contain the virus but they may also be a manifestation of structural racism itself. Indeed, one of the defining characteristics of structural racism is its invisibility to the dominant racial/ethnic group [20]. The situation also highlights the need for more comprehensive data on the impact of the COVID-19 pandemic on racial and ethnic minorities, in particular for those living in disadvantaged neighborhoods.

In that regard, eminent public health researchers have raised the alarm regarding missing or incomplete data on racial/ethnic minorities [14,21]. Indeed, data on the impact of COVID-19 on racial and ethnic minorities were not systemically and uniformly collected across the United States [22]. This resulted in the skewing of our understanding of the deadly evolution of the virus within these communities, while hampering the ability to provide them with timely and adapted public health interventions and care [14,23]. Although US state and local public health departments were required—at latest by August 1, 2020—to report demographic data for COVID-19 cases, Krieger and colleagues [23] have shown that, despite the new reporting requirements, compliance was far from being achieved and much work still needs to be done in this regard.

In the COVID-19 era, the underreporting and inadequate reporting of racial and ethnic information create data gaps that hamper the proper functioning of public health institutions in initiating culturally appropriate measures to prevent the spread of the virus in these communities [24]. Given the fast pace at which the pandemic is evolving and the continuous need for updated public health responses, a paramount question is how these data gaps and structural racism could influence public health practice, in particular when innovative data-driven approaches are being considered as complementary to traditional public health interventions. In this paper, we thus reflect on how structural racism in the health care/public health domain and the defective collection of data on racial and ethnic minorities could undermine the health equity and social justice goals of precision public health (PPH) interventions.

8.3 Structural Racism and Defective Data: Potential Impact on PPH

To respond to this worrying public health situation while making use of existing and emerging data sources, Rasmussen and colleagues [25] argue for the need to implement PPH interventions as an additional tool to fight the pandemic. PPH means “the application and combination of

new and existing technologies […] to tailor preventive interventions for at-risk groups and improve the overall health of the population" [26]. Horton [27] is even more explicit in highlighting the importance of data in PPH and characterizes it as being "about using the best available data to target more effectively and efficiently interventions of all kinds to those most in need." At least two distinct approaches to PPH exist. The first one is a reductionist version, where PPH focuses solely on the use of genetic information to tailor interventions to specific subgroups of the population (with the risk of neglecting foundational considerations of public health such as the impact of the social determinants of health). The second (wider and more encompassing) version does not limit itself to genetic information, but also considers other sources of data (eg, big data, granular population surveillance data, and data from mobile apps) to guide public health practice [25,28]. Given the limits of focusing solely on genetic data to guide interventions (eg, risk of exacerbating existing health inequalities [29]), we find the second version of PPH more promising, as it uses a plethora of data sources [28], in particular if it will ally both high-risk strategies and population-based approaches to maximize the impact of public health interventions [30].

Although still in its infancy [31], PPH has already shown promise in the fight against COVID-19. For instance, at the start of the pandemic, pathogen genomics such as whole genome sequencing analysis, coupled with epidemiological data, was successfully used in the Netherlands to not only monitor the emergence of local or regional clusters of SARS-CoV-2, but also to help in understanding transmission patterns while guiding public health interventions in breaking the chain of transmission of SARS-CoV-2 [32]. Additionally, some innovative sources of digital data are also being used to reduce the transmission of SARS-CoV-2. The usefulness of participatory disease surveillance systems has already been demonstrated for the early detection of other transmissible diseases such as influenza-like illness [33]. In such systems, citizens can actively get involved in the public health response by directly reporting COVID-19–related symptoms via mobile apps or digital platforms [34]. Furthermore, the use of COVID-19 contact trackers, whereby cellphone tracking data are used to alert people who might have been in contact with an active case of COVID-19 [35], is another example of how PPH can be helpful in a pandemic context.

Given the disproportionate burden of the pandemic on racial/ethnic minorities, as previously argued, it would then be legitimate to consider them as part of those "most in need" [27] and thus principal supposed beneficiaries of PPH interventions during the COVID-19 pandemic. However, given the aforementioned issues related to structural racism and the skewed

collection of data on minorities, it is important to consider the following two issues before the widespread deployment of PPH interventions in connection to COVID-19. First, one must reflect on how structural racism can influence the generation and use of data sets in public health practice (eg, data racism) [36]. Second, it is critical to explore how the use of certain data-driven technologies—which are becoming essential components of PPH—could lead to novel racial/ethnic discrimination in public health interventions for those "most in need" [27].

One set of technologies that has the capacity to both improve or worsen health inequities between ethnic and racial groups if employed in clinical care and PPH interventions is machine learning [37]. Machine learning can be defined as "a branch of artificial intelligence (AI) focused on building applications that learn from data and improve their accuracy over time without being programmed to do so" [38]. There are 3 main classifications of machine learning approaches, namely supervised, semisupervised, and unsupervised learning, depending on whether the machine learning algorithms are trained on labeled, semilabeled, or unlabeled data sets, respectively [39]. If such technologies are used in PPH, one of the important aspects to consider is the appropriateness of data sets used to train machine learning algorithms, in particular if they are to be used in a population that is either underrepresented in the training data sets or has long been systemically disadvantaged and marginalized [40,41]. Indeed, individual and societal biases can be encoded in big data and other training data sets destined for public health practice and medical care [37,40]. This phenomenon is sometimes described as data racism, a term that refers to "the multiple systems and technologies - deployed in a range of fields - that either primarily target or disproportionately impact migrants and people of [color]" [36]. To better explain how data racism—combined with structural racism—could impact PPH interventions through machine learning techniques in the future, one can look at the unfortunate experiences emerging in the field of data-driven predictive policing [42].

PPH and predictive policing can be compared since they function according to analogous principles. Indeed, one of the foreseeable goals of PPH is to forecast disease outbreaks and identify hotspots or subgroups of the population for tailored interventions based on big data predictive analytics [43]. Similarly, predictive policing aims at forecasting the likelihood of a crime being committed in a specific location, to then prioritize focused police interventions in certain at-risk areas (eg, by having more frequent police patrols) or on people having some prespecified characteristics deemed relevant by the used software. Existing predictive policing software companies, such as PredPol, aspire to also be fair, since the "starting point is data: objective, agreed-upon facts that can be used to guide the discussion" [44,45]. PredPol claimed

to provide fairer and more objective risk evaluation and predictions regarding crimes than subjective police assessment [44]. However, Richardson and colleagues [42] argue that, although PredPol took significant actions to reduce bias in their data sets when training their machine learning algorithms (eg, by excluding traffic citation data and data on drug-related crimes), such measures still do not capture the whole complexity and diversity of police interactions where bias can be introduced into the data. They also highlighted the methodological difficulties for vendors of such technologies to identify "these problematic practices and policies in real-time; therefore, any system that includes recent or live data may be subject to additional undocumented biases" [42]. Consequently, the alleged promise of being fair cannot always be kept, since data sets on which predictive policing is based often present relevant fallacies.

We can reasonably expect that PPH interventions might encounter an analogous set of problems, in particular considering the previously discussed COVID-19–related data crisis for racial and ethnic minority groups. Indeed, within the precincts of structural/data racism, machine learning algorithms will likely replicate some degree of discrimination unless appropriate measures are taken to address the situation [40]. Therefore, it is important for machine learning developers to have an adequate understanding of structural racism and its potential real-world ramifications through their software [46]. This could help ensure that the developed machine learning algorithms both advance public health utility and promote a fair distribution of resources along racial and ethnic lines, while minimizing the risks of worsening health inequities [46]. However, it is also important to note that addressing the technical flaws of machine learning algorithms is catering only to the downstream consequences of structural racism and therefore this cannot be the silver bullet to reduce health inequities between racial and ethnic groups. To bridge the disparities between racial and ethnic groups will likely require changes at the societal, institutional, and individual levels, so that the upstream influences of structural racism are mitigated [47]. A few additional potential solutions that can help in tackling algorithmic biases have been discussed in a previous publication [40].

Aside from technological considerations, it is also paramount to tackle the low representation of racial/ethnic minorities (in particular those of African descent) in AI and machine learning communities. This can also be considered a consequence of structural racism in industrial and academic settings [48]. Racial and ethnic diversity in these communities could help safeguard against blatant and implicit discrimination toward minorities, even if these minority researchers could themselves be subsequently exposed to consequences from the power structures in place.

Indeed, it is documented that some of the effects of structural racism in the workplace also involve microaggressions toward these minorities (eg, harassment, disrespect, racial slurs) by their White colleagues, which undermine their capacity to work, while operationalizing the corporate culture of maintaining racial hierarchy [20,49].

There are also other means by which structural racism could impact the generation of data for PPH during the COVID-19 pandemic. One of them is through the racial and ethnic digital divide, whereby minorities, as a consequence of their often-lower socioeconomic status, would be less likely to engage in PPH activities due to poorer internet access [50]. According to the Pew Research Center, in 2019, White individuals still have better access to the internet in comparison to other racial/ethnic groups (92% versus 86% in Hispanic individuals and 85% in African Americans) [51]. Therefore, public health surveillance systems relying on internet-collected data (eg, social media mining to guide interventions against COVID-19) may be particularly vulnerable to the underrepresentation of minority groups in the gathered data sets. The problem might also be that of overrepresentation of minorities in internet-collected data. For example, a recent study has found that—despite worse internet access—ethnic/racial minority groups are more likely to post COVID-19–related information on social media, possibly due to a "reversal of digital divides" or the fact that the pandemic has disproportionately impacted racial and ethnic minority groups, and that social media is often perceived as a coping strategy for stress and for obtaining community support [50]. Either way, the issue remains that proper public health surveillance should ensure the ethnic and racial representativeness of the collected data before initiating any public health intervention [50], or at least perform some sort of data adjustment to account for the selection bias.

Lastly, it cannot be ignored that structural racism could lead to more biased data sets for ethnic and racial minority groups because of the influence of sociopolitical factors on the functioning of public health institutions [52]. It is important to note that institutional (or structural) racism can influence the functioning of institutions in a number of ways (ie, at individual and organizational levels). For instance, institutions can either become increasingly racialized and adopt racist motivations (eg, excluding or limiting employment of qualified minorities in decision-making positions, or employing a few of them only, as part of racial capitalism [53], to showcase that they are dedicated to diversity) or they can be burdened by other organizational structures that impede their achievements in terms of racial and health equity (eg, lack of resources and low care given to people of color) [52]. Indeed, it is well known that hospitals in marginalized neighborhoods (eg, those with a high percentage of people of color) are

underfunded, understaffed, and sometimes filled with less skilled health care professionals, which then limit their ability to offer services and treatment comparable to those offered in other neighborhoods [15,54].

Therefore, such public health institutions could be hampered in generating high quality data on racial and ethnic minority groups due to (1) the limited resources dedicated to these groups, (2) the unrepresentative racial and ethnic composition of their decision-making teams and health care professionals, leading to viewing racial disparities from a "White framing," (3) the lack of adequately trained professionals to fulfill these duties, and (4) hospital segregation—that is, the refusal of some of the most resourceful hospitals to treat racial/ethnic minority groups suffering from COVID-19 due to their lower socioeconomic conditions (eg, inferior health insurance plans, which is also a known consequence of structural racism) [40,54,55]. The unfortunate consequence could be that of contributing to the generation of biased data sets that do not reflect the lived reality of racial and ethnic minorities in facing COVID-19. Therefore, PPH interventions relying on these data sets would likely be limited in their effectiveness and health equity could never be achieved for all racial and ethnic groups, unless these different parameters are given due consideration.

8.4 Conclusions

In this article, we have highlighted the role of structural racism and the presence of defective data for racial and ethnic minorities as important factors to consider in designing ethically acceptable public health policies during and after the COVID-19 pandemic. Moreover, we have discussed PPH, an innovative approach to tackle public health issues such as COVID-19, which can however encounter several ethical challenges if the aforementioned issues of structural racism and defective data collection are not tackled. Our aim is not that of discouraging the use of PPH, which we consider as an important new element in the public health toolbox that deserves to be fully implemented to fight the COVID-19 pandemic; rather, we want to highlight a few issues that need to be considered by policy makers and scientists developing PPH measures to make sure that their efforts to improve public health do not ignore the danger posed by structural racism and defective data on minorities. In this regard, we join our voices to those cited in this article on the importance of having improved, harmonized, and nationwide data collection systems that are as free as possible from the influences of structural racism and inclusive of all racial and ethnic groups. Although insufficient on their own to promote health equity among racial and ethnic groups, these measures could be an important contribution to the effective and nondiscriminatory use of PPH approaches that are inclusive of the racial and

ethnic composition of the societies in which they are deployed. PPH approaches deserve to be better planned and their effectiveness critically assessed, and this will not be achieved unless their data and technological foundations are deep-rooted in health equity and social justice.

Abbreviations

AI: artificial intelligence

PPH: precision public health

Acknowledgments

This research was funded by the Swiss National Science Foundation (NRP 74 "Smarter health care," project "Advancing SMart solutions and eliminating barriers for the Acquisition, Analysis, and Sharing of Health data in Switzerland [SMAASH]," grant 407440_167356). The publication fee was also funded by the Swiss National Science Foundation. The funder had no role in the writing of the manuscript or in the decision to publish the results.

Authors' Contributions

LDG was involved in the conceptualization of the manuscript, wrote the original draft and subsequently reviewed and edited the manuscript following other coauthors' suggestions and modifications made to the manuscript. AM was involved in the conceptualization of the manuscript, and in reviewing and editing of the original draft and subsequent versions. TW was involved in the conceptualization of the manuscript, supervised LDG during the writing of the original draft, and reviewed and edited the original draft and subsequent versions. BSE was involved in the conceptualization of the manuscript, and reviewed and edited the original draft and subsequent versions. BSE was also responsible for project administration and acquired funding from the Swiss National Science Foundation for this publication. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Swiss National Science Foundation, grant number 407440_167356. The APC was also funded by the Swiss National Science Foundation.

Conflicts of Interest

None declared.

8.5 References

1. 11 Terms You Should Know to Better Understand Structural Racism. The Aspen Institute. 2016. URL: https://www.aspeninstitute.org/blog-posts/structural-racism-definition/ [accessed 2022-02-25]
2. COVID-19 Racial and Ethnic Health Disparities 2020. Centers for Disease Control and Prevention. 2020 Dec. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-disparities/index.html [accessed 2022-02-25]
3. Yearby R, Mohapatra S. Law, structural racism, and the COVID-19 pandemic. J Law Biosci 2020;7(1):lsaa036
4. The color of coronavirus: COVID-19 deaths by race and ethnicity in the US. American Public Media Research Lab. URL: https://www.apmresearchlab.org/covid/deaths-by-race [accessed 2022-02-25]
5. Risk for COVID-19 Infection, Hospitalization, and Death By Race/Ethnicity. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html#footnote03 [accessed 2022-02-25]
6. Poteat T, Millett GA, Nelson LE, Beyrer C. Understanding COVID-19 risks and vulnerabilities among black communities in America: the lethal force of syndemics. Ann Epidemiol 2020 Jul;47:1-3 [FREE Full text] [CrossRef] [Medline]
7. Williams RA. Historical Perspectives of Healthcare Disparities. In: Williams RA, editor. Eliminating Healthcare Disparities in America: Beyond the IOM Report. Totowa, NJ: Humana Press; 2007:3-19.
8. Hall WJ, Chapman MV, Lee KM, Merino YM, Thomas TW, Payne BK, et al. Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. Am J Public Health 2015 Dec;105(12):e60-e76.
9. Farmer B. The Coronavirus Doesn't Discriminate, But U.S. Health Care Showing Familiar Biases. NPR. 2020 Apr 02. URL: https://www.npr.org/sections/health-shots/2020/04/02/825730141/the-coronavirus-doesnt-discriminate-but-u-s-health-care-showing-familiar-biases [accessed 2022-02-25]
10. Williams DR, Cooper LA. COVID-19 and Health Equity-A New Kind of 'Herd Immunity'. JAMA 2020 Jun 23;323(24):2478-2480.
11. Mulchan SS, Wakefield EO, Santos M. What COVID-19 Teaches Us About Implicit Bias in Pediatric Health Care. J Pediatr Psychol 2021 Feb 19;46(2):138-143

12. Bibbins-Domingo K, Kushel M. Op-Ed: How the pandemic divided California in two. Los Angeles Times. 2020 Aug 16. URL: https://www.latimes.com/opinion/story/2020-08-16/coronavirus-cases-california-inequality [accessed 2022-02-25]
13. Yearby R. Structural Racism and Health Disparities: Reconfiguring the Social Determinants of Health Framework to Include the Root Cause. J Law Med Ethics 2020 Sep;48(3):518-526.
14. Krieger N. ENOUGH: COVID-19, Structural Racism, Police Brutality, Plutocracy, Climate Change-and Time for Health Justice, Democratic Governance, and an Equitable, Sustainable Future. Am J Public Health 2020 Nov;110(11):1620-1623.
15. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. Lancet 2017 Apr 08;389(10077):1453-1463.
16. Richardson J, Mitchell B, Meier H, Lynch E, Edlebi J. The Lasting Impact of Historic 'Redlining' on Neighborhood Health: Higher Prevalence of COVID-19 Risk Factors. National Community Reinvestment Coalition. 2020. URL: https://ncrc.org/holc-health/ [accessed 2022-02-25]
17. Martin CA, Jenkins DR, Minhas JS, Gray LJ, Tang J, Williams C, Leicester COVID-19 consortium. Socio-demographic heterogeneity in the prevalence of COVID-19 during lockdown is associated with ethnicity and household size: Results from an observational cohort study. EClinicalMedicine 2020 Aug;25:100466
18. Anand S, Montez-Rath M, Han J, Bozeman J, Kerschmann R, Beyer P, et al. Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study. Lancet 2020 Oct 24;396(10259):1335-1344.
19. Friston K, Costello A, Pillay D. 'Dark matter', second waves and epidemiological modelling. BMJ Glob Health 2020 Dec;5(12):e003978
20. Wiecek WM, Hamilton JL. Beyond the Civil Rights Act of 1964: Confronting Structural Racism in the Workplace. Louisiana Law Review. 2014. URL: https://digitalcommons.law.lsu.edu/lalrev/vol74/iss4/5/ [accessed 2022-02-25]
21. Dowling MK, Kelly RL. Policy Solutions for Reversing the Color-blind Public Health Response to COVID-19 in the US. JAMA 2020 Jul 21;324(3):229-230.
22. Gravlee CC. Systemic racism, chronic health inequities, and COVID-19: A syndemic in the making? Am J Hum Biol 2020 Sep;32(5):e23482
23. Krieger N, Testa C, Hanage WP, Chen JT. US racial and ethnic data for COVID-19 cases: still missing in action. Lancet 2020 Nov 07;396(10261):e81

24. Wade L. COVID-19 data on Native Americans is ‘a national disgrace.’ This scientist is fighting to be counted. Science. 2020 Sep 25. URL: https://www.science.org/content/article/covid-19-data-native-americans-national-disgrace-scientist-fighting-be-counted [accessed 2022-02-25]
25. Rasmussen SA, Khoury MJ, Del Rio C. Precision Public Health as a Key Tool in the COVID-19 Response. JAMA 2020 Sep 08;324(10):933-934.
26. Weeramanthri TS, Dawkins HJS, Baynam G, Bellgard M, Gudes O, Semmens JB. Editorial: Precision Public Health. Front Public Health 2018;6:121
27. Horton R. Offline: In defence of precision public health. Lancet 2018 Oct 27;392(10157):1504.
28. Chowkwanyun M, Bayer R, Galea S. 'Precision' Public Health - Between Novelty and Hype. N Engl J Med 2018 Oct 11;379(15):1398-1400.
29. Taylor-Robinson D, Kee F. Precision public health-the Emperor's new clothes. Int J Epidemiol 2019 Feb 01;48(1):1-6
30. Rose G. Sick individuals and sick populations. Int J Epidemiol 2001 Jun;30(3):427-32; discussion 433.
31. Khoury MJ, Holt KE. The impact of genomics on precision public health: beyond the pandemic. Genome Med 2021 Apr 23;13(1):67.
32. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, Dutch-Covid-19 response team. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat Med 2020 Sep;26(9):1405-1410.
33. Geneviève LD, Martani A, Wangmo T, Paolotti D, Koppeschaar C, Kjelsø C, et al. Participatory Disease Surveillance Systems: Ethical Framework. J Med Internet Res 2019 May 23;21(5):e12273
34. Garg S, Bhatnagar N, Gangadharan N. A Case for Participatory Disease Surveillance of the COVID-19 Pandemic in India. JMIR Public Health Surveill 2020 Apr 16;6(2):e18795.
35. Servick K. Cellphone tracking could help stem the spread of coronavirus. Is privacy the price? Science. 2020 Mar 23. URL: https://www.science.org/content/article/cellphone-tracking-could-help-stem-spread-coronavirus-privacy-price [accessed 2022-02-25]
36. Chander S. Data racism: a new frontier. European Network Against Racism. URL: https://www.enar-eu.org/Data-racism-a-new-frontier [accessed 2022-02-25]

37. Khoury MJ, Armstrong GL, Bunnell RE, Cyril J, Iademarco MF. The intersection of genomics and big data with public health: Opportunities for precision public health. PLoS Med 2020 Oct;17(10):e1003373

38. Machine Learning. IBM. URL: https://www.ibm.com/cloud/learn/machine-learning [accessed 2022-02-25]

39. El Naqa I, Murphy MJ. What Is Machine Learning? In: El Naqa I, Li R, Murphy MJ, editors. Machine Learning in Radiation Oncology. Cham: Springer International Publishing; 2015.

40. Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. BMC Med Ethics 2020 Feb 19;21(1):17.

41. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep;25(9):1337-1340.

42. Richardson R, Schultz J, Crawford K. Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. New York University Law Review. 2019. URL: https://www.nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/ [accessed 2022-02-25]

43. Khoury MJ, Engelgau M, Chambers DA, Mensah GA. Beyond Public Health Genomics: Can Big Data and Predictive Analytics Deliver Precision Public Health? Public Health Genomics 2018;21(5-6):244-250

44. Are We at a Tipping Point in Police-Community Relations? PredPol. 2020 Jun 11. URL: https://blog.predpol.com/are-we-at-a-tipping-point-in-police-community-relations [accessed 2022-02-25]

45. PredPol. URL: https://www.predpol.com/ [accessed 2022-02-25]

46. Robinson WR, Renson A, Naimi AI. Teaching yourself about structural racism will improve your machine learning. Biostatistics 2020 Apr 01;21(2):339-344

47. Wachter S, Mittelstadt B, Russell C. Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law. SSRN Journal 2021;123(3):3792772.

48. Editorial. Technology can't fix this. Nat Mach Intell 2020 Jul 15;2(7):363-363.

49. Roscigno VJ, Garcia LM, Bobbitt-Zeher D. Social Closure and Processes of Race/Sex Employment Discrimination. The Annals of the American Academy of Political and Social Science 2016 Sep 08;609(1):16-48.

50. Campos-Castillo C, Laestadius LI. Racial and Ethnic Digital Divides in Posting COVID-19 Content on Social Media Among US Adults: Secondary Survey Analysis. J Med Internet Res 2020 Jul 03;22(7):e20472

51. Internet/Broadband Fact Sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2022-02-25]

52. Griffith DM, Childs EL, Eng E, Jeffries V. Racism in organizations: The case of a county public health department. J Community Psychol 2007;35(3):287-302.

53. Leong N. Racial Capitalism. SSRN Journal 2012:2009877.

54. Glastris P, Longman P. American hospitals are still segregated. That's killing people of color. The Washington Post. 2020 Aug 05. URL: https://www.washingtonpost.com/outlook/2020/08/05/segregated-hospitals-killing-people/ [accessed 2022-02-25]

55. Feagin J, Bennefield Z. Systemic racism and U.S. health care. Soc Sci Med 2014 Feb;103:7-14.

Chapter 9: Structural racism in precision medicine: leaving no one behind

Lester Darryl Geneviève[1], Andrea Martani[1], David Shaw[1,2], Bernice Simone Elger[1,3] and Tenzin Wangmo[1]

[1]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland.

[2]Care and Public Health Research Institute, Maastricht University, Maastricht, the Netherlands.

[3]University Center of Legal Medicine, University of Geneva, Geneva, Switzerland.

Citation: Geneviève, L.D., Martani, A., Shaw, D., Elger, B.S., Wangmo, T. Structural racism in precision medicine: leaving no one behind. BMC Med Ethics 21, 17 (2020). DOI:10.1186/s12910-020-0457-8 [CC BY 4.0]

"I often tell my students not to be misled by the name "artificial intelligence" – there is nothing artificial about it. AI is made by humans, intended to behave by humans, and ultimately, to impact human lives and human society" – Fei-Fei Li

9.1 Abstract

Background: Precision medicine (PM) is an emerging approach to individualized care. It aims to help physicians better comprehend and predict the needs of their patients while effectively adopting in a timely manner the most suitable treatment by promoting the sharing of health data and the implementation of learning healthcare systems. Alongside its promises, PM also entails the risk of exacerbating healthcare inequalities, in particular between ethnoracial groups. One often-neglected underlying reason why this might happen is the impact of structural racism on PM initiatives. Raising awareness as to how structural racism can influence PM initiatives is paramount to avoid that PM ends up reproducing the pre-existing health inequalities between different ethnoracial groups and contributing to the loss of trust in healthcare by minority groups.

Main body: We analyse three nodes of a process flow where structural racism can affect PM's implementation. These are: (i) the collection of biased health data during the initial encounter of minority groups with the healthcare system and researchers, (ii) the integration of biased health data for minority groups in PM initiatives and (iii) the influence of structural racism on the deliverables of PM initiatives for minority groups. We underscore that underappreciation of structural racism by stakeholders involved in the PM ecosystem can be at odds with the ambition of ensuring social and racial justice. Potential specific actions related to the analysed nodes are then formulated to help ensure that PM truly adheres to the goal of leaving no one behind, as endorsed by member states of the United Nations for the 2030 Agenda for Sustainable Development.

Conclusion: Structural racism has been entrenched in our societies for centuries and it would be naïve to believe that its impacts will not spill over in the era of PM. PM initiatives need to pay special attention to the discriminatory and harmful impacts that structural racism could have on minority groups involved in their respective projects. It is only by acknowledging and discussing the existence of implicit racial biases and trust issues in healthcare and research domains that proper interventions to remedy them can be implemented.

9.2 Background

The Precision Medicine Initiative (PMI) working group defines PM as "an approach to disease treatment and prevention that seeks to maximize effectiveness by taking into account individual variability in genes, environment, and lifestyle" [1]. Indeed, technological advances and increasing computing power allow to do a more in-depth characterization of individuals' variability and hence their predisposition to diseases by considering not only their genomic

profiles but other factors, including other omics (e.g. metabolomics), and their environmental and mobile data. The goal of PM is to advance medical and scientific discoveries while offering more tailored, precise and accurate health interventions, which will maximize the health benefits for patients [2, 3]. With such an approach, individual well-being is monitored pro-actively, that is, PM is predictive, personalized, preventive and participatory (in this context, the terms PM, personalized medicine and "P4 medicine" are used interchangeably) [4, 5].

The defining features and goals of PM seem to make it complementary to the scope of health equity as defined by the World Health Organization [6] that "ideally everyone should have a fair opportunity to attain their full health potential and that no one should be disadvantaged from achieving this potential". In fact, PM is aimed at providing patients with preventive and therapeutic interventions based on their individual needs (e.g. their susceptibility profile to some diseases). On a practical level, the objective of PM initiatives is to help physicians better comprehend and predict the needs of their patients, so that they may adopt the most suitable treatment in a timely manner. This goal is promoted by sharing health data and implementing learning healthcare systems [7]. Given the overall continuity of the clinical objectives, PM should be considered an "evolutionary" rather than a "revolutionary" approach to clinical trials, medicine and clinical care [8]. Indeed, technological advances over the years (e.g. relatively cheap genomic sequencing or tumour profiling) are accelerating scientific discoveries and subsequent market approval of new therapeutics in comparison to conventional means (e.g. by reducing the required number of participants for clinical trials or even the need for a control group) [8].

Alongside its promises, PM also entails the risk of exacerbating health inequalities, in particular between racial and ethnic groups. The fact that PM has a participatory component requires that different racial and ethnic groups trust and actively engage in PM initiatives [9], which is, however, extremely challenging. Minority communities often face discrimination in healthcare and receive poor medical treatment [10]. Outreach to these communities – especially in the research field – has also been characterized by a long history of exploitation, abuse and marginalization [11]. Events like the Tuskegee syphilis experiment [12] or cases like that of Henrietta Lacks [13] are often cited as causes of distrust by minority groups towards healthcare services and involvement in research projects. However, the relatively lower participation rate of minority groups in health research is not simply a matter of distrust and unwillingness [14]. In a review of enrolment decisions of more than 70,000 individuals for participation in health research, Wendler and colleagues [14] have shown that willingness to participate did not differ

significantly between ethno-racial groups and argued that underrepresentation of minority populations is more likely due to the research design of the single study or to limited accessibility. Aside from its cause, lower participation of minority groups has also contributed to most genetic databases used for research purposes containing data on participants of predominantly European ancestry [11, 15]. From an analysis of Genome-Wide Association Studies (GWAS) representing 1.7 million samples conducted in 2009, it resulted that 96% of participants were of European ancestry. Seven years later, the same GWAS analysis revealed that racial and ethnic representativeness of the samples still had a long way to go. In spite of the colossal 35 million samples collected, 81% of participants were still of European ancestry [16]. That racial and ethnic minorities are marginalized in the research field has also been underscored by the authors who concluded that "the message being broadcast by the scientific and medical genomics community to the rest of the world is currently a harmful and misleading one: the genomes of European descendants matter the most" [16].

The situation in the healthcare sector is similarly discouraging: a prominent scholar has recently underscored that, with respect to health, "America is Failing its Black Mothers" [17]. For instance, the pregnancy-related mortality ratio in the US during 2011–2016 for black women was 42.4 deaths per 100,000 live births, more than three times higher than for white women [18]. In a report by Amnesty International [19], titled "Deadly Delivery, the maternal health care crisis in the USA", it was reported that some healthcare providers do not take into account the timely healthcare needs of women of colour and treat them suboptimally or sometimes even try to dissuade them from seeking medical care, which left these women feeling "ignored or treated with disdain by staff". Therefore, health disparities and pregnancy outcomes of women of colour are influenced by systemic factors that either regulate access to healthcare or influence the quality of care offered to minority groups.

Although much of the research at the intersection of healthcare and race is conducted in the United States [20], the situation of minority groups is likely to be similar in Europe [10]. For instance, a large European study on end-stage renal disease demonstrated that "black and Asian patients were about half as likely to receive a kidney transplant as white patients, a finding that was not explained by differences in cause of kidney failure" [21]. The authors highlighted that disparities, both in terms of mortality on renal replacement therapy and decreased access to renal transplantation, could not be explained completely by the cause of kidney failure, and that other factors such as the socioeconomic, cultural, environmental and even other biological factors were likely to be involved [21]. Moreover, it is worrying that little is known on the

epidemiological profile of minority groups in European high-income countries due to their often unfortunate exclusion in epidemiological studies [22]. Their inclusion would contribute to a better understanding of the healthcare inequalities faced by members of these communities. Additionally, this lack of epidemiological data is a missed opportunity in the era of PM. For instance, coupling epidemiological data with genetic data could provide some additional insights on how socio-cultural, economic and environmental factors influence biological pathways in minority groups and contribute to the pathogenesis of certain diseases (e.g. heart disease) [23].

As rightly pointed out by Bayer and Galea [24], PM initiatives tend to focus mostly on individual health rather than considering how social determinants and structural realities (e.g. residential segregation for minority groups) have shaped and are continuing to shape population health. This, coupled with the pre-existing structural problems illustrated above, supports discriminatory actions against minority groups, which altogether increases their vulnerability to adverse health outcomes [19], provoking distrust in the healthcare system, which will precipitate to PM initiatives [25]. Therefore, it is crucial to identify and better understand the underlying systemic factors that jeopardize the trust of minority groups in healthcare professionals, and institutions now dedicated in advancing the goals of PM. Without a trusting relationship between minority groups and PM initiatives, these are unlikely to succeed in their research objectives, as representative collection and integration of health data (from EHRs, tissue samples, etc.) will be compromised [25].

One societal phenomenon that can in part explain such disparities in the quality of healthcare provided to different ethnic and racial groups is racism [26]. According to *Oxford English Dictionary*, the word "racism" is defined as "prejudice, discrimination, or antagonism directed against someone of a different race based on the belief that one's own race is superior" or "the belief that all members of each race possess characteristics, abilities, or qualities specific to that race, especially so as to distinguish it as inferior or superior to another race or races" [27]. Given the influence that racism still has in healthcare, the marginalization of racial and ethnic minorities might not be the only reason why PM does not stand up to its promise of providing equal chances for all. From our perspective, an underestimated factor whereby PM can contribute to inequality in healthcare and research between different ethnoracial groups is its susceptibility to racism in general and to structural racism in particular. The term *structural racism* refers to "ideologies, practices, processes, and institutions that operate at the macro level to produce and reproduce differential access to power and to life opportunities along racial and

ethnic lines" [28]. Over centuries, it has been entrenched in numerous countries, influencing the way medicine is taught and practiced as well as the functioning of healthcare institutions [29]. This might help to understand why – although genetic predisposition or unhealthy lifestyle, biological inferiority, socio-economic factors, and medical distrust are put forward as some of the reasons contributing to the persistence of healthcare inequalities between ethnic and racial groups [30, 31] – even when some of such factors are taken into account, these inequalities remain [32]. Furthermore, unhealthy lifestyle and socio-economic factors themselves are in turn partly a product of structural racism and discrimination [30].

Claiming that PM initiatives might be subject to the influence of structural racism might sound controversial, since – theoretically – PM endorses social and racial justice between racial and ethnic groups. For instance, the majority of PM initiatives are implemented with the aim of ensuring ethnic diversity and appropriate ethnoracial representation in their cohorts (e.g. All of Us Research Program, New York University's Human Project and Project Baseline) [33]. Such measures aspire at ascertaining that no racial or ethnic group is left behind and that every individual, irrespective of his or her racial and ethnic backgrounds, benefits from advances in healthcare. However, in spite of its ambition to promote social and racial justice, PM might nonetheless end up accentuating healthcare inequalities between different racial and ethnic groups if it covertly adopts the existing cultural processes such as identification (racialization, e.g. associating racial stereotypes with some therapeutic options, and stigmatization) and rationalization of health services provided to certain racial groups and ethnic minorities [34, 35]. Therefore, PM might reiterate the current *status quo* in healthcare, where very few racial groups are privileged to the detriment of others, especially if structural racism is not taken into account.

In this debate paper, we discuss the ways in which the implementation of PM might be particularly vulnerable to structural racism in healthcare and research, and forecast its potential impacts in the upcoming era of PM. Specifically, we analyse three nodes in the process flow of PM where structural racism can have an impact. These nodes are part of a process flow: (1) collection of biased health data during the initial encounter of patients with healthcare system and researchers, (2) integration of biased health data for PM initiatives and (3) the influence of structural racism on deliverables of PM initiatives. After analysing the interaction of PM and structural racism, we propose future actions to help make PM initiatives truly adhere to the goal of "Leaving no one behind", as endorsed by member states of the United Nations for the 2030 Agenda for Sustainable Development [36].

9.3 Main text

The three nodes of structural racism in precision medicine

In the ecosystem of PM, there seem to be three specific nodes where structural racism can have an impact: the quality of health data collected; the integration of these data in PM initiatives; and the development of new therapeutics, diagnostics or disease prevention strategies. In this context, the concepts of clinical and research data are grouped together under health data, as boundaries between clinical practice and research are blurring in the convergence framework of PM, learning healthcare systems and implementation science [37]. A learning healthcare system would allow the improvement of care over time by collecting data in the clinical encounter and using them to improve the effectiveness and efficiency of current clinical practice, by facilitating the exchange of information between clinical and research sectors [9, 37]. It would thus act as a bridge for the integration of new discoveries made through PM initiatives into routine clinical care, whereas implementation science would be the "catalyst" to such transition by providing strategies to promote the operationalisation of PM initiatives' new findings [37, 38]. These three nodes are used in the following sections to structure our arguments.

First node: collection of biased health data during initial encounter with healthcare system and researchers

The first node depicts the initial encounter between minority populations and healthcare providers and/or researchers, which leads to the production of biased health data, collected (among others) in electronic health records (EHR), biobanks or different research data banks [9]. For the first node, there are two distinct aspects, which need to be carefully considered. Firstly, minority groups are under-represented in current health services and research datasets, due to unequal access to healthcare and clinical studies [14]. Such underrepresentation can negatively affect the quality of health services provided to their members, since they might be treated according to guidelines informed by biased data – in the form of data that disproportionately represent people of the majority ethnic or racial group [33]. For example, PM is spearheading the fight against certain types of cancer [2], owing to technological advances made in genomics (e.g. with the advent of next-generation sequencing allowing the identification of a huge number of variants [39]), cancer biology and other relevant fields. It thus provides a more molecular-based and individualized approach to dealing with both primary and recurrent/metastatic tumours [40]. However, there are numerous barriers hindering the participation of minority groups to genetic testing for evaluating cancer risk [41]. Without enough genetic data for some minority populations, it will be almost impossible to distinguish

pathological from benign variants in these subgroups, and consequently, evaluating their risks of developing a certain cancer type might be compromised. Therefore, minority groups at high-risk of developing a disease will not benefit from high quality disease preventive measures [31], even if granted access to similar a treatment to that offered to the majority group.

Secondly, minority groups are more susceptible to receive suboptimal care due to implicit provider bias in healthcare, which also feeds into the problem of biased health data. Indeed, it is known that healthcare providers, irrespective of their specialization fields or levels of experience, suffer from an implicit racial or ethnic bias when dealing with people of colour [42]. Such bias negatively affects their ability to provide efficient health services to minority groups, due to uncontrolled thoughts or feelings that influence their clinical judgement [42, 43]. For instance, a 2016 US study showed that medical students and residents held false beliefs concerning biological differences between black and white people, which negatively influenced their assessment of pain and treatment recommendations in people of colour [44]. Similarly, another study showed that black children were prescribed fewer antibiotics than their white counterparts when examined by the same physician [45]. In the same vein, a study found that healthcare providers in US emergency departments have a high implicit preference for non-Hispanic white people over the American Indian community [46]. Hence, it is clear that in healthcare systems where most professionals are of Caucasian origin, people of colour are at risk of not being given equal access and a level of care comparable to that offered to patients of Caucasian origin. Given these premises, it is probable that even if new individualized treatments are available to people of colour, PM initiatives will fall short of their goals.

Indeed, equal access to individualized prevention and treatment might be compromised by unconscious racial bias already existing in the healthcare context. Due to these negative implicit racial stereotypes [47], healthcare providers might not prescribe new therapeutic drugs to these communities or might treat them suboptimally. Moreover, this “aversive racism” (i.e. having a high degree of implicit bias and a relatively low degree of explicit bias) during medical encounters, which is underrecognized and habitually unintended, leads minority groups to respond more negatively to physicians [48]. Aversive racism thereby undermines patient’s trust due to lower perceived quality of care, poor doctor-patient communication as well as a loss of interest in joint decision-making [48, 49]. These are all detrimental to the goals of PM initiatives, since they lead to biased data being produced for minority groups, who, in turn, are less likely to engage in research activities. Although biased data originate predominantly from subjective interpretation (e.g. biased clinical evaluation by physicians) rather than objective

measurements (e.g. results of an MRI scan, blood tests), it nonetheless remains possible that objective data on minority groups are not being captured optimally due to biased clinical evaluation of their medical conditions.

Since PM initiatives gather data from both new and existing sources (e.g. electronic health records, biobanks, etc.), Ferryman and Pitcan [33] highlighted that, in the era of PM, "it is important to recognize the potential limitations within these data today that come from historical legacies of bias and discrimination". On top of that, we argue that it is also important for PM initiatives to better understand the limitations of new data collected by physicians or researchers involved in PM initiatives that belong to the majority group. Indeed, the iterative nature of PM initiatives and learning healthcare systems implies that data are gathered continuously to generate new insights into individuals' health, which are thereafter implemented in practice for better-individualized prevention and treatment. However, if minority groups suffer from racially discriminatory actions in clinical practice and are offered less effective healthcare interventions due to biased nonrepresentative data, the chain of healthcare improvements based on reliable routinely collected clinical data may be compromised from its very start.

Due to past betrayals of trust, minority groups might be reluctant to engage with their healthcare system, leading not only to a lack of interests in PM activities but also limited data representativeness from these groups [33]. It is crucial to understand the expectations and fears of minority groups regarding their participation in PM initiatives. For instance, a recent study showed that minority groups also fear that, by participating to PM initiatives, results of these initiatives might unwillingly contribute to further racial discrimination from either the healthcare system (e.g. being denied access to treatment because it is specific to an ethnic/racial group) or by their healthcare insurers and employers (e.g. loss of employment opportunities or higher insurance premiums) [25]. This altogether has negative repercussions on the quality of health data collection efforts to provide evidence-based healthcare and on the development of accurate clinical guidelines or treatments for these communities [33].

The production of biased health data for minority groups leads us to the second node of the process flow.

Second node: integration of biased health data for PM initiatives

The second node characterizes the integration of biased health data from minority populations into PM initiatives, leading to their faulty interpretation and thus to misuse in scientific research and clinical practice [33]. With respect to this node, it is crucial to acknowledge the twofold

potential damage resulting from biased PM initiatives for minority groups. Firstly, health data have always been prone to historical biases and minority groups are already paying a high price for them. For instance, current clinical guidelines are largely developed from cohorts of white men, whose risks factors for developing a particular disease could be very different from men (and women) belonging to minority populations. Therefore, this sampling bias implies that the threshold required to justify certain medical interventions or disease prevention strategies would differ based on an individual's racial or ethnic background [33]. One concrete example that was widely covered in the literature is the *Framingham Coronary Heart Disease Risk Functions*, a risk assessment score used for the primary prevention of coronary heart disease (CHD). The *Framingham Risk Score* was originally developed from a population of principally white cohorts in the USA to predict the risk of CHD and subsequent appropriate preventive measures. It was shown to overestimate the risks of cardiovascular diseases not only in some minority groups (e.g. Hispanic and Japanese American men [50]), but also in some European (e.g. Germany [51]) and nonEuropean countries (e.g. China [52]), thereby highlighting the need for recalibration. Until eligibility for interventions and interventions themselves are calibrated, it is thus probable that data required for the good functioning of PM initiatives are not being captured for minority groups. Secondly, the former problem (historical bias), coupled with structural racism, could be amplified with the increasing use of artificial intelligence (AI) technologies to assist physicians and researchers in their routine work [33].

The application of AI technologies is rapidly increasing in the healthcare sector [53], and according to Ferryman and Pitman [33], AI is also a necessary feature of PM due to the increasing availability of big health data sources. Indeed, AI technologies are considered to be one of the solutions to help researchers and physicians interpret the ever-growing amount of health data produced on a daily basis, which already greatly exceed physicians' analytical capabilities [53]. However, there is increasing concern that these AI technologies are hugely dependent on the data that they are trained with and can subsequently aggravate societal biases present in the training databases [54]. Whereas decisions by healthcare providers or researchers might be only *intermittently* influenced by racial bias, decisions made by machine learning algorithms will be *systematically* biased every time the latter are used, leading to more discrimination against minority groups and to a much larger scale [55]. Indeed, how historical bias in the training datasets and hence in algorithmic decisions can lead to more discrimination is perfectly illustrated by the case of the AI tool named COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) used in the US judicial sector. COMPAS was a software designed to support judicial decision-making concerning potential recidivism of

offenders. It assigned probability scores to defendants on whether they were likely to break the law within 2 years after being released from prison. COMPAS was shown to be biased against black offenders due to presumed historical bias in the data, an element which led to more black people being kept in prison rather than being released just because of their ethnicity [56]. This case was particularly interesting as one could even argue that the bias of the software against black people was not immediately perceived by the judges using the AI tool because racial and ethnic prejudices are so deeply rooted and implicit that they easily go unnoticed. In the same vein, it would not be surprising if historical bias in health datasets used in the training of AI technologies for PM initiatives, coupled with structural racism, ended up reproducing existing healthcare inequalities between racial and ethnic groups. If so, physicians – just like judges in the case of COMPAS – would be very unlikely to identify flawed medical decision-making induced by AI, because of their pre-existing prejudices.

On top of biased medical decision-making, AI may also have a negative impact on the recruitment of people of colour in clinical trials. Clinical trials have traditionally been known to be time- and resource-consuming, with difficulties in "matching the right trial with the right patient", but AI has been forecasted to provide the solution to this by automating the whole clinical trial matching through available health data sources [57]. For instance, DEEP 6 AI is a software company based in the US that analyses both structured and unstructured data using machine learning, natural language processing and medical ontologies with the aim of matching eligible patients to potential clinical trials in a timely manner [58]. Another example comes from Microsoft, who, as part of their Microsoft Healthcare Bot initiative [59], use machine reading to assign suitable patients to clinical trials with the aim of streamlining the whole recruitment process [60]. Similar to biased medical decision-making, we argue that the use of AI technologies in automatically assigning patients to clinical trials in the PM era may also be negatively influenced by historical bias in the health datasets (e.g. EHRs) [33] and by structural racism. If not properly designed, these AI technologies could exacerbate health inequalities between minority and majority groups by either excluding or limiting the eligibility of people of colour to participate in certain studies. For instance, Obermeyer and Mullainathan [61] discovered how an algorithm used in US healthcare on over 70 million patients was racially biased against black people. The algorithm reduced the chances of black people being enrolled in the "care management program", and the culprit was not the training datasets per se but rather the inappropriate choice of labels (e.g. healthcare costs) which did not provide the full picture regarding the health of black people [61]. Another example came from the University of Chicago hospital system, where researchers found that if postal codes had been used in their

machine-learning algorithm to optimize hospital resources, resources available for black people would have been diverted towards “wealthy white people, exacerbating existing biases in the system” [62]. As residential segregation is also a known consequence of structural racism [63], this shows how structural racism can have many repercussions on algorithmic decisions in the healthcare system.

Third node: influence of structural racism on deliverables of PM initiatives

The third node refers to the uptake of new disease prevention strategies, diagnostics and therapeutics from PM initiatives into the cycle of learning healthcare frameworks. The goal of a learning healthcare system is to provide better care to individuals over time by continuously collecting clinical encounter data and using them to develop strategies to improve the quality of care offered to patients. It thus provides a unique opportunity for findings of PM initiatives to be implemented in the routine clinical life cycle [37]. With respect to this node, the risk of racial discrimination is due to the potentially discriminatory effects of feeding biased data into a learning healthcare framework, especially because current healthcare systems are already designed and built for patients of the majority group [10] and are consequently not customized for minorities in terms of their reduced access to care. There are numerous reasons for reduced access to care, including (1) the fact that minority groups are sometimes unable to pay for health services due to lower health insurance coverage, (2) medical distrust as a result of previous racially discriminatory actions or perceived racism (which in itself is an additional detrimental stressor to the health of minority groups [26]) that delay or prevent access for treatment, and (3) the geographical variation in healthcare quality offered to minority groups [64–66]. In a learning healthcare framework, reduced access to healthcare implies the loss of important clinical encounter data from minority groups due to reduced or delayed contacts with physicians [65], which would normally help to improve the monitoring of disease evolution and subsequent appropriate treatment options. This bias induced through “invisibility” - caused by insufficient data or incomplete datasets on minority groups - can potentially lead to adverse and discriminatory health outcomes as easily as overtly flawed data [33].

Over the past few years, there has been a real commitment from pharmaceutical companies to advance the goals set by PM initiatives by producing new personalized medicines. In 2018, the Center for Drug Evaluation and Research of the FDA approved a record number of 25 new personalized therapeutics, which represented 42% of the total number of drug approvals for that year [67]. In this respect, another important aspect, in terms of minority groups’ access to adequate healthcare, may be the lack of interest in developing new therapeutic options for diseases more prevalent in minority groups, due to structural racism embedded in the world of

research and in drug development. From our perspective, Farooq and Strouse [68] gave an excellent example of potential racial bias in research and drug development by comparing two distinct diseases, each one predominantly affecting a different racial group.

Cystic fibrosis, affecting predominantly white populations, is an autosomal recessive disease resulting from a defect in the gene encoding for the chloride channel, CFTR (cystic fibrosis transmembrane conductance regulator). This alteration leads to pulmonary complications such as chronic bacterial infections, bronchiectasis and pulmonary fibrosis [69]. *Sickle-cell disease* (SCD), on the contrary, mainly affects people of colour. Globally, SCD is one of the most severe blood conditions, caused by a mutation in the beta globin gene, which leads to the production of sickle globin instead of beta globin, a component necessary for the production of normal haemoglobin. This genetic mutation causes the occlusion of blood vessels and haemolytic anaemia, resulting in many complications such as premature death, acute chest syndrome or cerebrovascular disease (e.g. stroke) [70]. Although both conditions have similar disease severity, and a lower percentage of patients in the United States suffers from *cystic fibrosis* in comparison to SCD, Farooq and Strouse [68] demonstrated that there are wide disparities in how funding was allocated for research by the National Institutes of Health (NIH) and private Foundations to study the two diseases. Moreover, research productivity in terms of PubMed-indexed articles and drug approvals were significantly higher for *cystic fibrosis* than for SCD, in spite that both diseases have similar numbers of clinical trials [68].

With reference to the last five progress reports (2014–2018) of the *Personalized Medicine Coalition* [67, 71– 74], we observed the same trend regarding FDA approval of personalized drugs for the two diseases. Two personalized drugs for *cystic fibrosis* were approved in this time span: Orkambi (*ivacaftor* and l*umacaftor*; 2015) and Symdeko (*ivacaftor* and *tezacaftor*; 2018); but none for SCD. Therefore, it remains imperative to underline that although minority groups could actively engage in PM initiatives by voluntarily contributing their data for research, the research and pharmaceutical sector might be biased in improving or finding new diagnostics and therapeutics for diseases prevalent in the white population. Such discrimination is partly caused by financial interests of pharmaceutical companies, which prioritize drug development for western market, as the countries can afford the high prices of developed drugs. In 2014, the ex CEO of Bayer, Marijn Dekkers, raised a lot of controversy when he declared that his company only produced cancer medication for "western patients who can afford it" and not for the Indian market, a statement condemned massively by *Médecins Sans Frontières* [75]. Therefore, if PM wants to achieve its equity goals and thereby safeguards the trust and long-

term engagement of minority groups, it is paramount to ensure that members of minority groups see the clear benefits that their communities will get in return for their participation to PM activities [25].

Connecting the nodes – some future actions

It is to be expected that the mentioned deleterious effects of structural racism will be reinforced over time in PM initiatives, due to the iterative data–exchange process between research and the clinical sector. Therefore, it is paramount to consider the impacts of structural racism at the very outset of PM initiatives, in order to prevent continued discriminatory treatment of minority groups in research and during clinical care. It is also important to recognize that trust [76] and engagement of minority groups in PM initiatives need to be safeguarded for PM to achieve its full potential. In the previous sections, we explored the nodes where structural racism could affect the implementation of PM initiatives, and forecasted that bias induced by structural racism in health datasets can have cascading deleterious effects on the health of minority groups. In the following sections, we recommend some potential actions that can help ensure those negative effects are mitigated (Fig. 1).

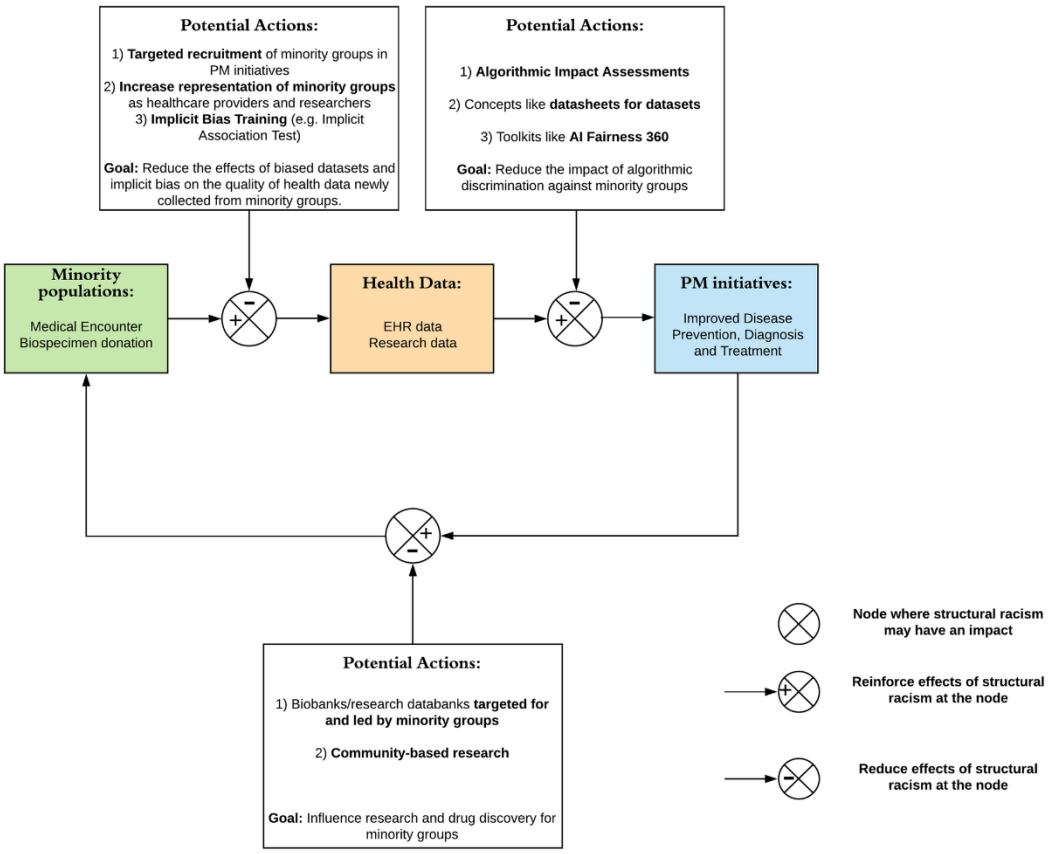


Fig. 1 Potential actions to reduce the cascading effects of structural racism on the quality of health data collection, integration and deliverables in precision medicine initiatives

Potential actions to reduce the impact of structural racism at node 1

Regarding the limited representation of minority groups in current health and research databases, some PM initiatives are already taking actions to address these issues. For instance, the *All of Us Research Program* [77] has been prioritizing minority groups for the collection of biospecimens and physical examinations. In a special report published in *NEJM*, the investigators stated that, as from July 2019, "more than 80% of these participants [over 175,000 participants donated their biospecimens] are from groups that have been historically underrepresented in biomedical research" [77]. This represents a huge step forward towards improving data representativeness for minority groups and ensuring that they will be offered healthcare interventions that are adequately tailored to their 'real' needs and not extrapolated from non-representative data. In contrast, the *UK biobank*, a valuable resource for PM [78], with over 500,000 study participants recruited over 2006 to 2010 [79], has adopted a different approach. Although explicitly acknowledging the limited generalizability of their data to the general population of the United Kingdom [80], the *UK biobank* relies on the large cohort size to cater for selection bias and the resulting limited data representativeness for *accurately* assessing "exposure-disease relationships" [81]. In this regard, we could not agree more with the correspondence from Keyes and Westreich [82], where it was argued that inferences derived from large sample sizes can also be skewed (to the detriment of external validity) and such aspects should be "taken more seriously in the UK biobank and other large data resources". Therefore, we deem more appropriate to have better targeted recruitment and interventions, like those carried out by the *All of Us Research Program*, which would help reduce selection bias and limited data representativeness of minority groups in PM initiatives.

Aside from tackling the problem of representativeness at the institutional level, actions needs to be taken also at the individual and professional level. According to 2017 statistics, the great majority of physicians and surgeons in the United States were white, accounting for 69.8% of the workforce, followed by Asian (21.1%), black (5.8%) and other minority groups [83]. As put forward by Cohan in a recent *NEJM* article [84], "… health care is not safe for people of colour as long as the overwhelming majority of U.S. physicians are white and we avoid examining where racism lives within us and how it lives through us". Therefore, we can reasonably argue that one of the reasons underlying the normalization of discriminatory actions against people of colour in healthcare is the lack of racial and ethnic representativeness in healthcare professions and the under-recognition of the impacts of structural racism by white physicians. The EU parliament has recently (26/03/2019) passed a resolution where it invites other "European institutions to adopt a workforce diversity and inclusion strategy that

establishes a strategic plan for the participation of ethnic and racial minorities in their workforce" [85]. This commitment may help to address this problem in Europe. However, it is important not only to increase the percentage of minority groups as physicians, researchers and in other healthcare professions, but also to request white physicians to see their routine clinical work in a new light when dealing with people of colour. A good option to assess their degree of implicit racial bias against minority groups is through the *Implicit Association Test* developed by Project Implicit [86], which aims to educate the population about unconscious biases. Data from Project Implicit have already been used to reveal how racial prejudice negatively influences birth outcomes for black women in the United States [87], or even the pervasiveness of implicit prejudices against the lesbian, gay, bisexual, and transgender (LGBT) community among healthcare providers [88]. Such tests could help white physicians to better identify and subsequently question the biased choice of treatment for minority groups.

These different measures will help to ensure that hospitals, research institutions and other similar structures function on principles, values and foundations which are representative of the ethnic and racial make-up of their society [89]. These actions will also contribute to improving the quality of health data collected on minority groups, since they will hopefully reduce racial discriminatory actions and restore trust. Gaining the trust of minority groups and ensuring that data collection is less affected by bias introduced by structural racism will forward the promised health benefits of PM initiatives to help bridge the health gaps between racial and ethnic groups.

Potential actions to reduce the impact of structural racism at node 2

The issues surrounding biased algorithmic decisions has not left lawmakers indifferent [90], in particular following recent notable big tech scandals (e.g. the Cambridge Analytica Affair [91] or the fact that the Department of Housing and Urban Development in the US is suing Facebook over discrimination in housing-related advertising [92]). To tackle some of these technology-related issues, a new bill has been recently introduced in both Houses of the US Congress, the *Algorithmic Accountability Act 2019*, which aims at ensuring fairer and nondiscriminatory algorithmic decisions. Although representing an important step in the fight against algorithmic discrimination, it has been underscored that this bill seems to be lacking in three important aspects: (1) at the level of enforcement, it relies on the Federal Trade Commission, which, as an agency, rarely enforces its settlements with privacy violators; (2) at the level of impact assessments, it lacks an avenue for diverse public participation, in particular from affected communities; and (3), also at the level of impact assessments, it does not provide for them to be made public [90]. One solution to these issues could come from the implementation of *Algorithmic Impact Assessments* (AIA) for public agencies, to ensure that *automated decision*

systems are not only assessed by involved stakeholders, but also by members of communities affected by these systems [93]. Within the AIA framework, the concerned agency would need to publicly disclose its definition of *automated decision system*, any assessments and external reviews made on the potential impacts of the system before its procurement, and the public would then be allowed to comment on the system and clarify its concerns with the agency. Additionally, the government would have the duty to ensure that the rights of affected individuals are respected by providing improved due process tools, in cases where an agency has not corrected a biased system. Such measures would hold the concerned agency accountable while safeguarding against unlawful discrimination or the non-respect of rights of affected communities [93].

Another solution put forward by Gebru and colleagues [54] is the concept of *datasheets for datasets*, which would help tackle the issues surrounding biases in training datasets for machine learning communities. According to the authors, each dataset should be accompanied by a datasheet explaining the characteristics of the dataset (e.g. motivation, composition, collection process, etc.). These datasheets could potentially address the biases in training datasets for machine learning processes by increasing not only transparency but also accountability within machine learning communities [54]. Researchers, tech companies, and physicians would thus be able to make a more informed choice in the selection of adequate datasets for a given task and therefore reduce the impact of biases against minority groups.

Assessment and corrective measures can be taken against algorithmic discrimination with either the training dataset, the learning procedure (i.e. the classifier) or the predictions of the AI tool. In this regard, IBM has proposed the *AI Fairness 360* (AIF360), an open source toolkit aimed to "promote a deeper understanding of fairness metrics and mitigation techniques; to enable an open common platform for fairness researchers and industry practitioners to share and benchmark their algorithms; and to help facilitate the transition of fairness research algorithms to use in an industrial setting" [94]. Depending on where the intervention is needed to avoid algorithmic bias in the AI cycle, AIF360 proposes three approaches, namely *pre-processing* (actions needed on the training dataset), *in-processing* (actions needed on the classifier) and *post-processing* (actions needed to correct predictions) bias mitigation algorithms [94]. Regardless of the instrument used, education on strategies to check and mitigate algorithmic bias in their tools could be extremely beneficial for AI developers active in the field of PM.

Potential actions to reduce the impact of structural racism at node 3

Another factor that might undermine the good health intentions of PM initiatives towards minority groups is the limited access to healthcare and new therapeutics. A first fundamental step to try remedy this situation is to intervene in the processes of creation and development of biobanks. According to Shaw and colleagues [95], "a biobank is any collection of human biological samples and linked data that is to be used for research". These, together with databanks, are globally regarded as essential research infrastructures for PM, allowing the collection of health data from large cohorts, and deriving "wisdom from crowds" to deliver individualized treatment [78]. However, in biobanks and databanks, there is often an underrepresentation of minority groups. This is not only the result of recruitment difficulties but also of the deliberate exclusion of these groups by scientists, as their inclusion in studies will lead to confounding results due to genetic variation [96]. The unfortunate consequence of such exclusion or underrepresentation is the exacerbation of healthcare inequalities between racial and ethnic groups, because it is more unlikely that treatments tailored to their needs are discovered.

To tackle this issue, efforts made to introduce biobanks specific to and led by minority groups should be praised and strongly encouraged. One such example is the BRAICELET project (Bio-Repository for American Indian Capacity, Education, Law, Economics and Technology), which aims to reduce health inequalities "through the establishment of a first-of-its-kind American Indian Biobank" [97]. In the BRAICELET project, American Indian communities are allowed to "lead collaborations with universities and research institutes across the nation to find culturally and real-time solutions to issues of disparity affecting American Indian communities" [97], enabling the implementation of programs that are tailored to the needs of these indigenous communities. Similarly, the National Institutes of Health and the Wellcome Trust jointly funded a large-scale initiative, called *Human Heredity and Health in Africa* (H3Africa) to allow the implementation of PM in the continent. H3Africa seeks to facilitate research on diseases affecting African populations by gathering genetic and environmental data on tens of thousands of participants [98]. The data gathered by H3Africa will be used to influence research in the field of pharmacogenomics, where African communities have long been marginalized, with the goal of discovering drugs most susceptible to benefit the health of African populations [98]. Some minority groups also view community-based research not only as being more valuable to their communities, but also as a means of motivating them to participate in activities of PM initiatives [25].

9.4 Conclusions

Structural racism has been entrenched in our societies for centuries and it would be naïve to believe that its impacts will not spill over in the era of PM. In this perspective, PM initiatives around the world should pay particular attention to the potential impacts that structural racism could have on their respective projects, and consider the three nodes analysed in this paper. PM initiatives should embrace the responsibility to mitigate the described impacts of structural racism, in particular those impacts upon which they have direct control. Therefore, careful consideration needs to be given to the choice of health datasets used in their projects to limit racial biases (e.g. the *datasheet for datasets* concept can be a good starting point) and their collaborators (e.g. physicians, researchers and technology developers) need to be better informed about the detrimental and insidious impacts of structural racism on their activities. For instance, the *Implicit Association Test* could allow physicians to reflect upon their routine clinical practice to identify situations where their attitudes and medical decisions for minority groups might have been influenced by unconscious biases and promptly try to remedy the situation by sensitizing themselves to the cultural values and perspectives of minority groups. These initiatives should also encourage the implementation of specific biobanks and other research databanks targeted for minority groups, with the mandatory inclusion of members of these communities at the management level, to ensure that scientific discoveries are stirred towards improving or finding new treatment for diseases affecting predominantly minority groups (e.g. through community-based research). Although not falling directly under their control, PM initiatives should also encourage and lobby for an adequate representation of ethnic minorities in healthcare professions so that the quality of health data collected for minority groups is improved, with the aim of reducing healthcare inequalities between racial and ethnic groups.

Above all, we believe that it is only by openly acknowledging and discussing the existence of implicit racial biases and trust issues in the healthcare and research domains that proper interventions can be implemented against structural racism. PM could offer a unique opportunity to bridge some of the long-standing racial gaps in healthcare and research. It, however, requires that the deleterious impacts of structural racism are carefully considered and addressed during the implementation of PM initiatives. This will help to prevent the reproduction and perpetuation of the current healthcare inequalities between different ethno-racial groups.

Abbreviations

AI: Artificial intelligence;

AIA: Algorithmic Impact Assessments;

BRAICELET: Bio-Repository for American Indian Capacity, Education, Law, Economics and Technology;

COMPAS: Correctional Offender Management Profiling for Alternative Sanctions;

EHR: Electronic Health Record;

GWAS: Genome-Wide Association Studies;

H3Africa: Human Heredity and Health in Africa;

LGBT: lesbian, gay, bisexual, and transgender;

PM: Precision Medicine;

SCD: Sickle-cell disease

Acknowledgements
The authors acknowledge the financial support provided by the Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356). The views expressed in this article are those of the authors and not necessarily those of the funder. LDG thanks Georg Starke for the insightful discussions and comments on the manuscript.

Authors' contributions
LDG, AM, DS, BSE, TW participated to the conceptualization of the manuscript. LDG wrote the first draft with the help of TW. LDG, AM, DS, BSE, TW contributed to the writing, editing, and critical evaluation of the manuscript. The authors approved the submission of the final version of the manuscript.

Funding
Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356). The funder played no role in the writing of the manuscript.

Availability of data and materials
Not applicable.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

9.5 References

1. PMI Working Group. The Precision Medicine Initiative Cohort Program Building a Research Foundation for 21st Century Medicine [Internet]. National Institutes of Health; 2015. [cited 2019 November 18]. Available from: https://www.nih.gov/sites/default/files/research-training/initiatives/ pmi/pmi-working-group-report-20150917-2.pdf.

2. Collins FS, Varmus H. A new initiative on precision Medicine. N Engl J Med. 2015;372(9):793–5.

3. Ashley EA. Towards precision medicine. Nat Rev Genet. 2016;17:507.

4. Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. New Biotechnol. 2012;29(6):613–24.

5. Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. BMC Med. 2018;16(1):150.

6. WHO. Health equity [Internet]. World Health Organization; 2019. [cited 2019 November 18]. Available from: https://www.who.int/topics/health_ equity/en/.

7. Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. Nature. 2015;526(7573):336–42.

8. Rubin R. A precision Medicine approach to clinical trials. JAMA. 2016;316(19): 1953–5.

9. Ginsburg GS, Phillips KA. Precision Medicine: from science to value. Health affairs (Project Hope). 2018;37(5):694–701.

10. Bhopal RS. Racism in health and health care in Europe: reality or mirage? Eur J Pub Health. 2007;17(3):238–41.

11. Cohn EG, Henderson GE, Appelbaum PS. Distributive justice, diversity, and inclusion in precision medicine: what will success look like? Genet Med. 2016;19:157.

12. Gamble VN. Under the shadow of Tuskegee: African Americans and health care. Am J Public Health. 1997;87(11):1773–8.

13. Lee SSJ, Cho MK, Kraft SA, Varsava N, Gillespie K, Ormond KE, et al. “I don’t want to be Henrietta lacks”: diverse patient perspectives on donating biospecimens for precision medicine research. Genet Med. 2019;21(1):107–13.

14. Wendler D, Kington R, Madans J, Van Wye G, Christ-Schmidt H, Pratt LA, et al. Are racial and ethnic minorities less willing to participate in health research? PLoS Med. 2006;3(2):e19-e.

15. Mersha TB, Abebe T. Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. Hum Genomics. 2015;9(1):1.

16. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nat News. 2016; 538(7624):161.

17. Roeder A. America is failing its black mothers [Internet]. Harvard Public Health; 2019. [cited 2019 April 8]. Available from: https://www.hsph.harvard. edu/magazine/magazine_article/america-is-failing-its-black-mothers/.

18. CDC. Pregnancy Mortality Surveillance System [Internet]. Centers for Disease Control and Prevention; 2020. [cited 2020 February 11]. Available from: https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pregnancymortality-surveillance-system.htm.

19. Amnesty International. Deadly delivery the maternal health care crisis in the USA. Amnesty International: United Kingdom; 2010.

20. Paradies Y. A systematic review of empirical research on self-reported racism and health. Int J Epidemiol. 2006;35(4):888–901.

21. Tjaden LA, Noordzij M, van Stralen KJ, Kuehni CE, Raes A, Cornelissen EA, et al. Racial disparities in access to and outcomes of kidney transplantation in children, adolescents, and young adults: results from the ESPN/ERA-EDTA (European Society of Pediatric Nephrology/European renal associationEuropean Dialysis and transplant association) registry. Am J Kidney Dis. 2016;67(2):293–301.

22. Stronks K, Snijder MB, Peters RJG, Prins M, Schene AH, Zwinderman AH. Unravelling the impact of ethnicity on health in Europe: the HELIUS study. BMC Public Health. 2013;13:402.

23. Mensah GA, Jaquish C, Srinivas P, Papanicolaou GJ, Wei GS, Redmond N, et al. Emerging concepts in precision Medicine and cardiovascular diseases in racial and ethnic minority populations. Circ Res. 2019;125(1):7–13.

24. Bayer R, Galea S. Public health in the precision-Medicine era. N Engl J Med. 2015;373(6):499–501.

25. Kraft SA, Cho MK, Gillespie K, Halley M, Varsava N, Ormond KE, et al. Beyond consent: building trusting relationships with diverse populations in precision Medicine research. Am J Bioeth. 2018;18(4):3–20.

26. Williams DR, Mohammed SA. Discrimination and racial disparities in health: evidence and needed research. J Behav Med. 2009;32(1):20–47.

27. Oxford English Dictionary. Racism [Internet]. Oxford University Press; 2019. [cited 2019 April 8]. Available from: https://en.oxforddictionaries.com/ definition/racism.

28. Viruell-Fuentes EA, Miranda PY, Abdulrahim S. More than culture: structural racism, intersectionality theory, and immigrant health. Soc Sci Med. 2012; 75(12):2099–106.

29. Bailey Z, Krieger N, Agénor M, Graves J, Linos N, Bassett M. Structural racism and health inequities in the USA: evidence and interventions. Lancet. 2017; 389(10077):1453–63.

30. Bassett MT, Graves JD. Uprooting institutionalized racism as public health practice. Am J Public Health. 2018;108(4):457–8.

31. Ramaswami R, Bayer R, Galea S. Precision Medicine from a public health perspective. Annu Rev Public Health. 2018;39:153–68.

32. Eastman P. IOM report: minorities receive unequal medical treatment even when access is similar. Oncol Times. 2002;24(5):31–2.

33. Ferryman K, Pitcan M. Fairness in Precision Medicine. Data & Society; 2018.

34. Hicken MT, Kravitz-Wirtz N, Durkee M, Jackson JS. Racial inequalities in health: Framing future research. Soc Sci Med. 2018;199:11–8.

35. Lamont M, Beljean S, Clair M. What is missing? Cultural processes and causal pathways to inequality. Soc Econ Rev. 2014;12(3):573–608.

36. United Nations. Leaving no one behind [Internet]. United Nations; 2016. [cited 2019 April 8]. Available from: https://unstats.un.org/sdgs/report/2016/ leaving-no-one-behind.

37. Chambers DA, Feero WG, Khoury MJ. Convergence of implementation science, precision Medicine, and the learning health care system: a new model for biomedical research. JAMA. 2016;315(18):1941–2.

38. Fisher ES, Shortell SM, Savitz LA. Implementation science: a potential catalyst for delivery system reform. JAMA. 2016;315(4):339–40.

39. Blumenthal GM, Mansfield E, Pazdur R. Next-generation sequencing in oncology in the era of precision Medicine. JAMA Oncol. 2016;2(1):13–4.

40. Morris LGT, Chandramohan R, West L, Zehir A, Chakravarty D, Pfister DG, et al. The molecular landscape of recurrent and metastatic head and neck cancers: insights from a precision oncology sequencing platform. JAMA Oncol. 2017;3(2):244–55.

41. Hann KEJ, Freeman M, Fraser L, Waller J, Sanderson SC, Rahman B, et al. Awareness, knowledge, perceptions, and attitudes towards genetic testing for cancer risk among ethnic minority groups: a systematic review. BMC Public Health. 2017;17(1):503.

42. Hall WJ, Chapman MV, Lee KM, Merino YM, Thomas TW, Payne BK, et al. Implicit racial/ethnic Bias among health care professionals and its influence on health care outcomes: a systematic review. Am J Public Health. 2015;105(12):e60–76.

43. van Ryn M, Burgess DJ, Dovidio JF, Phelan SM, Saha S, Malat J, et al. The impact of racism on clinician cognition, behavior, and clinical decision making. Du Bois Rev. 2011;8(1):199–218.

44. Hoffman KM, Trawalter S, Axt JR, Oliver MN. Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. Proc Natl Acad Sci. 2016;113(16):4296–301.

45. Gerber JS, Prasad PA, Localio AR, Fiks AG, Grundmeier RW, Bell LM, et al. Racial differences in antibiotic prescribing by primary care pediatricians. Pediatrics. 2013;131(4):677–84.

46. Puumala SE, Burgess KM, Kharbanda AB, Zook HG, Castille DM, Pickner WJ, et al. The role of Bias by emergency department providers in Care for American Indian Children. Med Care. 2016;54(6):562–9.

47. Penner LA, Blair IV, Albrecht TL, Dovidio JF. Reducing racial health care disparities: a social psychological analysis. Policy Insights Behav Brain Sci. 2014;1(1):204–12.

48. Penner LA, Dovidio JF, West TV, Gaertner SL, Albrecht TL, Dailey RK, et al. Aversive racism and medical interactions with black patients: a field study. J Exp Soc Psychol. 2010;46(2):436–40.

49. Williams DR, Wyatt R. Racial Bias in health care and health: challenges and opportunities. JAMA. 2015;314(6):555–6.

50. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. For the CHDRPG. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA. 2001;286(2):180–7.

51. Hense HW, Schulte H, Lowel H, Assmann G, Keil U. Framingham risk function overestimates risk of coronary heart disease in men and women from Germany--results from the MONICA Augsburg and the PROCAM cohorts. Eur Heart J. 2003;24(10):937–45.

52. Liu J, Hong Y, D'Agostino RB Sr, Wu Z, Wang W, Sun J, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese multi-provincial cohort study. JAMA. 2004; 291(21):2591–9.

53. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.

54. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumeé III H, et al. Datasheets for datasets. arXiv preprint arXiv:180309010. 2018.

55. Žliobaitė I. Measuring discrimination in algorithmic decision making. Data Min Knowl Disc. 2017;31(4):1060–89.

56. Courtland R. Bias detectives: the researchers striving to make algorithms fair. Nature. 2018;558(7710):357–60.

57. CB Insights. The Future Of Clinical Trials: How AI & Big Tech Could Make Drug Development Cheaper, Faster, & More Effective [Internet]. CB Insights; 2018. [cited 2019 May 6]. Available from: https://www.cbinsights.com/ research/clinical-trials-ai-tech-disruption/.

58. DEEP 6 AI. How it works [Internet]. DEEP 6 AI; 2019. [cited 2019 May 6]. Available from: https://deep6.ai/how-it-works/.

59. Microsoft. Microsoft Healthcare Bot [Internet]. Microsoft; 2019. [cited 2019 June 5]. Available from: https://www.microsoft.com/en-us/research/project/ health-bot/.

60. Bass D. Microsoft Built a Bot to Match Patients to Clinical Trials [Internet]. Bloomberg; 2019. [cited 2019 June 5]. Available from: https://www. bloomberg.com/news/articles/2019-03-01/microsoft-built-a-bot-to-matchpatients-to-clinical-trials.

61. Obermeyer Z, Mullainathan S. Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70 Million People. Proceedings of the Conference on Fairness, Accountability, and Transparency; 2019; Atlanta, GA, USA. New York: ACM; 2019.

62. Nordling L. A fairer way forward for AI in health care. Nature. 2019; 573(7775):S103–s5.

63. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Rep. 2001;116(5):404–16.

64. Bulatao RA, Anderson NB. Understanding racial and ethnic differences in health in late life: a research agenda: National Academies Press; 2004.

65. Rhee TG, Marottoli RA, Van Ness PH, Levy BR. Impact of perceived racism on healthcare access among older minority adults. Am J Prev Med. 2019;56(4): 580–5.

66. Gaskin DJ, Dinwiddie GY, Chan KS, McCleary R. Residential segregation and disparities in health care services utilization. Med Care Res Rev. 2012;69(2):158–75.

67. PMC. Personalized Medicine at FDA: A Progress & Outlook Report [Internet]. Personalized Medicine Coalition; 2019. [cited 2019 November 15]. Available from: http://www.personalizedmedicinecoalition.org/Resources/ Personalized_Medicine_at_FDA_An_Annual_Research_Report.

68. Farooq F, Strouse J. Disparities in foundation and Federal Support and development of new therapeutics for sickle cell disease and cystic fibrosis [abstract]. Blood. 2018;132(Suppl 1):4687.

69. O'Sullivan BP, Freedman SD. Cystic fibrosis. Lancet. 2009;373(9678):1891–904.

70. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. Lancet. 2010; 376(9757):2018–31.

71. PMC. Personalized Medicine at FDA: 2014 Progress Report [Internet]. Personalized Medicine Coalition. [cited 2019 November 15]. Available from: http://www.personalizedmedicinecoalition.org/Resources/Personalized_ Medicine_at_FDA_An_Annual_Research_Report.

72. PMC. Personalized Medicine at FDA: 2015 Progress Report [Internet]. Personalized Medicine Coalition. [cited 2019 November 15] . Available from: http://www.personalizedmedicinecoalition.org/Resources/Personalized_ Medicine_at_FDA_An_Annual_Research_Report.

73. PMC. Personalized Medicine at FDA: 2016 Progress Report [Internet]. Personalized Medicine Coalition. [cited 2019 November 15]. Available from: http://www.personalizedmedicinecoalition.org/Resources/Personalized_ Medicine_at_FDA_An_Annual_Research_Report.

74. PMC. Personalized Medicine at FDA: 2017 Progress Report [Internet]. Personalized Medicine Coalition. [cited 2019 November 15]. Available from: http://www.personalizedmedicinecoalition.org/Resources/Personalized_ Medicine_at_FDA_An_Annual_Research_Report

75. Balasegaram M. MSF response to Bayer CEO statement that medicines developed only for western patients [Internet]. Médecins Sans Frontières; 2014. [cited 2019 November 15]. Available from: https://msfaccess.org/msfresponse-bayer-ceo-statement-medicines-developed-only-western-patients

76. Adjekum A, Ienca M, Vayena E. What is trust? Ethics and risk governance in precision Medicine and predictive analytics. Omics. 2017;21(12):704–10.

77. The All of Us Research Program Investigators. The “all of us” research program. N Engl J Med. 2019;381(7):668–76.

78. UK Biobank data on 500,000 people paves way to precision medicine. Nature. 2018;562(7726):163–4.

79. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.

80. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol. 2017;186(9):1026–34.

81. UK Biobank. Access matter: representativeness of the UK Biobank resource [Internet]. UK Biobank; 2017. [cited 2019 November 15]. Available from:

http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-mattersrepresentativeness-1.pdf

82. Keyes KM, Westreich D. UK Biobank, big data, and the consequences of non-representativeness. Lancet. 2019;393(10178):1297.

83. DATA USA. Physicians & surgeons [Internet]. DATA USA. [cited 2020 February 11]. Available from: https://datausa.io/profile/soc/291060/#intro.

84. Cohan D. Racist like me - a call to self-reflection and action for white physicians. N Engl J Med. 2019;380(9):805–7.

85. The European Parliament. European Parliament resolution of 26 March 2019 on fundamental rights of people of African descent in Europe (2018/ 2899(RSP)) [Internet]. The European Parliament; 2019. [cited 2019 November 15]. Available from: http://www.europarl.europa.eu/doceo/document/TA-8-2 019-0239_EN.html.

86. Project Implicit. About us [Internet]. Project Implicit; 2011. [cited 2019 June 3]. Available from: https://www.projectimplicit.net/about.html

87. Orchard J, Price J. County-level racial prejudice and the black-white gap in infant health outcomes. Soc Sci Med. 2017;181:191–8.

88. Sabin JA, Riskind RG, Nosek BA. Health care Providers' Implicit and explicit attitudes toward lesbian women and gay men. Am J Public Health. 2015; 105(9):1831–41.

89. Graham L, Brown-Jeffy S, Aronson R, Stephens C. Critical race theory as theoretical framework and analysis tool for population health research. Crit Public Health. 2011;21(1):81–93.

90. Kaminski ME, Selbst AD. The Legislation That Targets the Racist Impacts of Tech [Internet]. The New York Times; 2019. [cited 2019 June 3]. Available from: https://www.nytimes.com/2019/05/07/opinion/tech-racismalgorithms.html

91. Schneble CO, Elger BS, Shaw D. The Cambridge Analytica affair and internet-mediated research. EMBO Rep. 2018;19(8):e46579.

92. Corbin K. HUD Is Suing Facebook For Housing Discrimination[Internet]. Forbes; 2019. [cited 2019 June 3]. Available from: https://www.forbes.com/sites/kennethcorbin/2019/03/28/hud-suing-facebook-for-housingdiscrimination/#183bc0797547

93. Reisman D, Schultz J, Crawford K, Whittaker M. Algorithmic impact assessments: a practical framework for public agency accountability [Internet]. AI NOW Institute; 2018. [cited 2019 June 5]. Available from: https://ainowinstitute.org/aiareport2018.pdf.

94. Bellamy R, Dey K, Hind M, Hoffman S, Houde S, Kannan K, et al. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM J Res Dev. 2019;63(4/5):4:1–4:15.

95. Shaw D, Elger BS, Colledge F. What is a biobank? Differing definitions among biobank stakeholders. Clin Genet. 2014;85(3):223–7.

96. Tutton R. Biobanks and the inclusion of racial/ethnic minorities. Race/ Ethnicity: Multidisciplinary Global Contexts; 2009. p. 75–95.

97. Stanford Medicine. BRAICELET: Bio-Repository for American Indian Capacity,

Education, Law, Economics and Technology [Internet]. Stanford Medicine; 2019. [cited 2019 April 11]. Available from: http://med.stanford.edu/sphere/ research-intiatives/braicelet.html

98. Mulder N, Abimiku A, Adebamowo SN, de Vries J, Matimba A, Olowoyo P, et al. H3Africa: current perspectives. Pharmacogenomics Pers Med. 2018;11: 59–66.

Chapter 10: General Discussion

"Being deeply knowledgeable on one subject narrows one's focus and increases confidence, but it also blurs dissenting views until they are no longer visible, thereby transforming data collection into bias confirmation and morphing self-deception into self-assurance." – Michael Shermer

10.1 Importance of thesis findings for precision health approaches

On the one hand, the empirical branch of this thesis includes four chapters: a systematic review, two interview-based studies, and a scoping review. The systematic review [1] is the first publication reporting comprehensively and comparatively on factors influencing the harmonization of health data and its processing by analyzing 345 Danish and Swiss studies. Given the extensive multi-stakeholder collaboration needed for PPH and PM to achieve their goals, this review provides insights on the different types of barriers that such initiatives are likely to encounter in the Swiss context due to its highly fragmented health data landscape. Thereafter, solutions adapted to the local context were formulated to remedy some of the identified barriers. Furthermore, Danish solutions to issues of harmonization and processing of health data were also evaluated and recommendations were made accordingly for their potential implementation in the Swiss health data landscape to try cater for its current weaknesses. Importantly, the review also showed that certain types of barriers (e.g. ethical, legal or socio-cultural, etc.) were underreported in the included studies, which could show either an under-appreciation of these barriers by the authors or it is the result of strategies used by the latter to bypass publication obstacles. However, such underreporting of these barriers is detrimental to solve issues of data harmonization and processing. In this regard, researchers might unwittingly and wrongly attribute additional weight to barriers that are more reported and tangible than others (e.g. those of a technical nature), but also that the underreporting of barriers represents a missed opportunity for other researchers and stakeholders to learn from them and plan accordingly their future research projects [1]. Lastly, the systematic review already had some policy impact, being cited in a report of the European Commission titled "Assessment of the EU Member States' rules on health data in the light of GDPR", which has for objective to underline "possible differences and identifying elements that might affect the cross-border exchange of health data in the EU, and examining the potential for EU level action to support health data use and re-use" [2].

The qualitative findings delineated in two chapters[15] were dedicated to investigating aspects of fair data sharing from the perspectives of expert Swiss stakeholders involved in the healthcare and research domains. Indeed, the concept of fair data sharing was investigated at two distinct levels. First, at a systemic level, where the hypercompetitive academic environment and legal uncertainty governing data sharing activities impose unfair conditions on researchers. Second,

[15] Geneviève LD, Martani A, Perneger T, Wangmo T, Elger BS. Systemic Fairness for Sharing Health Data: Perspectives From Swiss Stakeholders. Frontiers in Public Health. 2021;9.; Geneviève LD, Martani A, Elger BS, Wangmo T. Individual notions of fair data sharing from the perspectives of Swiss stakeholders. BMC Health Services Research. 2021;21(1):1007.

at an individual level, where notions of fair data sharing were operationalized through a distributive justice lens to identify fair or unfair data sharing practices used by stakeholders involved in the data-access negotiation process. These two studies are different but complementary to the majority of scientific publications since they explored the fairness of data sharing activities for the original data collectors and data recipients, not for data subjects. Indeed, the current trend has been up-to-now to foster data sharing as a fairness requirement for study participants or patients donating their data, and for the reproducibility of research findings, as discussed in chapters 4 and 5.

Regarding the scoping review [3], it is the first time a study has investigated how research ethics are implemented in the Influenzanet Consortium, a participatory disease surveillance platform where citizens volunteer to report ILI[16] symptoms via web-based and mobile technologies. Acknowledging the added value of such innovative surveillance systems in the fight against infectious diseases whilst highlighting the need for these systems to adapt to the evolving data protection landscape (e.g. the GDPR[17]), this study has shown that there is a need to harmonize and clarify ethical approaches adopted by each country-specific platform, in particular given that they all have similar data processing activities. This finding was important to promote future collaborations between members of the Influenzanet consortium by reducing the chances that such implementation disparities in research ethics are not the source of future obstacles to fruitful collaborations and effective public health surveillance [3], since epidemics – as the COVID-19 pandemic has proven – are increasingly not confined within national borders.

On the other hand, the theoretical branch of the thesis is comprised of three chapters, each tackling important and timely problematics related to the processing of health data in general or specifically in relation to data discrimination in PPH and PM, to foster racial equity in health.

The paper [4] titled "Participatory disease surveillance systems: ethical framework" was a thoughtful and interdisciplinary collaboration initiated with representatives of country-specific platforms of the Influenzanet Consortium. Recognizing the epistemological limitations in the scoping review since all details relevant to the operationalization of research ethics are not listed in included publications, the author undertook this multi-country collaboration to gain a deeper insight on ELSIs arising from digital public health surveillance and how they were managed in practice by directly involving and collaborating with co-authors from the country-specific platforms of Influenzanet. These platforms were those of Denmark, France, Italy, the

[16] Acronym for: Influenza-like illness
[17] Acronym for: General Data Protection Regulation

Netherlands, Switzerland, and the research collaboration also included the participation of AI[18] experts from the German Research Center for Artificial Intelligence (DFKI). The latter are involved in designing data collection and analysis tools for the Influenzanet platforms and their mobile applications. This is one of the few studies that provides an in-depth assessment of the ELSIs of such participatory platforms, whilst providing ethical recommendations that have been assessed and validated by the interdisciplinary group of experts involved. Therefore, this paper forwards an easily implementable and robust ethical framework, which is applicable to Influenzanet-like platforms in order to guide public health practice and research within ethically and legally acceptable margins whilst fostering public trust.

Another theoretical chapter[19] titled, "Precision Public Health and Structural Racism in the United States: Promoting Health Equity in the COVID-19 Pandemic Response", aims to highlight the ongoing data crisis and structural racism undermining the efficiency of public health interventions for racial and ethnic minorities, and how such crisis could lead to the exacerbation of health inequities through the combined use of PPH approaches and AI technologies such as machine learning. This paper is responding to the important societal call for more health equity and social justice, in particular following the widely mediatized police atrocities committed against members of disadvantaged and marginalized groups in not only the US (e.g. George Floyd and Breonna Taylor to name a few) but also in Switzerland (e.g. Mike Ben Peter, a black man who died from a heart attack in 2018 after being assaulted by the Police in Lausanne)[20].

The debate paper [5] titled, "Structural Racism in Precision Medicine: Leaving no one behind", was among the first scientific publications debating and forecasting the impact of structural racism in PM for racial and ethnic minorities from a data perspective. Given the paradigm shift occurring in the Swiss healthcare system to implement precision approaches to health and medicine (e.g. with the implementation of the SPHN [6]), this paper provides a unique approach to identifying weaknesses in the promised chain of healthcare improvement that PM and learning healthcare systems should bring. It analyzes three nodes of a process flow, from the collection and integration of health data to the elaboration of PM deliverables, where structural

[18] Acronym for: Artificial Intelligence

[19] Geneviève LD, Martani A, Wangmo T, Elger BS. Precision Public Health and Structural Racism in the United States: Promoting Health Equity in the COVID-19 Pandemic Response. JMIR Public Health Surveill. 2022;8(3):e33277.

[20] France24 (2020): "Switzerland grapples with its own 'George Floyd' case". URL: https://www.france24.com/en/20200616-switzerland-grapples-with-its-own-george-floyd-case. [Date accessed: March 3, 2021]

racism could have a significant impact. It also predicts that the deleterious effects of structural racism will be reinforced over time due to the iterative data exchange necessary between the research and healthcare domains, and the use of AI technologies such as machine learning [5]. Since its publication in 2020, it has been used as a discussion and educational material in several US universities, such as at the University of North Carolina School of Medicine[21], the Rosalind Franklin University of Medicine and Science[22], and cited by the Association of Black Cardiologists in their reply to the call from the chairman of the Committee on Ways and Means of the US House of Representatives. Indeed, the Honorable Richard E. Neal initiated a call for information to help address issues arising from the use of race in research and clinical algorithms[23]. Based on this publication, I was interviewed for a *Nature* article titled, "Health-care inequality could deepen with precision oncology"[24] where I shared my views on how structural racism could impact precision oncology for minorities. Moreover, delving into this topic of discrimination in PM and PPH has helped me in securing a research grant from the Käthe Zingg-Schwichtenberg Fund of the Swiss Academy of Medical Sciences, where I will start investigating the impact of structural racism in Swiss precision oncology programs, as of June 2021, using the debate paper as a blueprint[25].

10.2 Factors influencing the processing of health data in Switzerland

Chapters 3, 4 and 5 of this thesis provide not only important policy-making insights on barriers and facilitators influencing the sharing of health data in Switzerland, but also on those emerging when Swiss research projects go on a supranational level and need to share data with international partners.

[21] Carter-Edwards L (2020): "Pharmacogenetics, Personalized Medicine and Race: Understanding Implicit Bias". UNC, USA. URL: https://www.med.unc.edu/pharm/pharmacogenetics-personalized-medicine-and-race-understanding-implicit-bias-1/ [Date accessed: March 3, 2021]

[22] Rosalind Franklin University of Medicine and Science (2020): "CMS Health Disparities and Equity in Healthcare & Medical Education Journal Club- "Structural Racism in Precision Medicine: Leaving No One Behind"" – sponsored by Chicago Medical School, Office of Excellence in Diversity and Inclusion, Academic Learning Environment, and CMS Health Disparities and Equity in Medical Education Task Force. URL: https://www.rosalindfranklin.edu/about/strategic-initiatives/diversity-inclusion/calendar-of-events/ [Date accessed: March 3, 2021]

[23] Ofili EO, Ellis JL, Underwood PL, Badero OR and Breatheth K (2020). Ways and Means Committee. URL: https://waysandmeans.house.gov/sites/democrats.waysandmeans.house.gov/files/documents/Assoc%20of%20Black%20Cardiologists%20RFI%20Response%20to%20Chairman%20Neal%2011.30.20.pdf [Date accessed: March 3, 2021]

[24] Madhusoodanan J (2020). "Health-care inequality could deepen with precision oncology". Nature. URL: https://www.nature.com/articles/d41586-020-02678-7. [Date accessed: March 3, 2021]

[25] SAMW (2020). "Käthe Zingg-Schwichtenberg Fund". SAMW. URL: https://www.samw.ch/en/Funding/Kaethe-Zingg-Schwichtenberg-Fund.html. [Date accessed: March 3, 2021]

Identification and analysis of factors influencing the processing of health data

Through its systematic review [1], we learn that Swiss national and international projects in the health domain still encounter a large number of obstacles of varying natures and significance in comparison to their Danish counterparts.

First, technical barriers were frequently mentioned, in particular those related to heterogeneous data standards (i.e. semantics and structure), the low quality of collected datasets and the technical limitations in carrying out data sharing activities [1]. For instance, the inability for Swiss researchers to use the Swiss AHV[26] number as a universal personal identifier due to concerns about privacy risks [7], which would have otherwise facilitated the linkage of data on a specific individual from different databases in Switzerland. In comparison, Danish projects used extensively the universal personal identification number of Danish citizens to carry out data linkage activities. Moreover, barriers relating to data quality issues and data standards were also more equivalently reported in Swiss projects than in Danish ones, for which data quality issues were the main reported technical barrier [1]. The findings suggested that the highly fragmented Swiss healthcare system [8] certainly played a role in the more uniform reporting of these two types of technical barriers.

Indeed, the organization of healthcare in Switzerland is a federalist Bismarckian one [9]. The twenty-six federal states, known as *cantons*, enjoy a high degree of independence with regard to their respective financing policies and management of their healthcare systems, with little interference from Federal authorities [10], and where tens of different health insurers are active [11]. In addition, there is a clear demarcation between the providers of health services and health insurers [8, 12]. Such a diversity of healthcare actors – with sometimes different economic and political interests [13] – has contributed partly to the health data fragmentation in Switzerland [14]. Therefore, it is not surprising that issues related to differing data standards are still a major concern for the sharing of health data across cantons and institutions, particularly in multi-center Swiss projects [1].

Additionally, the systematic review revealed that, in comparison to Danish projects that were more retrospective in nature (often involving linking data between their nationwide disease registries), Swiss projects depended more on the prospective collection of health data. Such prospective data collection in Swiss projects was also shown in the review to necessitate more financial and motivational facilitators than Danish ones [1], since collecting data prospectively is often more expensive [15] and necessitates a greater effort from project developers. This

[26] Acronym for: Alters- und Hinterlassenenversicherung, which means Old Age and Survivors Insurance (OASI)

finding is also important because although prospective and retrospective studies have different strengths and limitations [15], it can be argued that each study type will likely raise a different set of ethical and legal barriers that will subsequently necessitate a different problem-solving approach by project developers.

In this regard, ethical and legal barriers were reported to a much lesser extent than technical ones in the systematic review [1], although these are some of the most important and pressing issues that need to be solved for the sharing and linkage of health data [16-18]. The main ethical and legal problematics identified in this systemic review were privacy issues, concerns about the respect for the autonomy of data subjects (e.g. difficulties in obtaining the informed consent of participants for certain studies), restrictive data protection regulations and lastly, heterogeneities in how country-specific legislations assessed data security and privacy aspects for multi-country research collaborations [1]. Regarding the latter, this is not only due to dissimilar territorial jurisdictions governing data sharing activities, but also through legal uncertainties around data sharing from both practical and conceptual perspectives [19].

In Switzerland, the "consent or anonymize" approach is still being used predominantly by project developers to solve ethical and legal issues [20]. Indeed, Swiss project developers usually tend to circumvent these problematics by either seeking informed consent from data subjects to allow the processing of their personal data or by fully anonymizing datasets, which allows their processing without consent [1]. Indeed, the resulting data are no longer considered as identifiable and therefore their processing does not fall under the remit of the Swiss Human Research Act (Art. 2) [21] or the GDPR (Recital 26) [22]. In comparison, our systematic review showed that Danish project developers have been supported in their data processing activities by regulations that allow processing of health data without REC approval or seeking consent from data subjects if these projects satisfied certain conditions. Furthermore, the Danish Data Protection Agency plays a more active role in data-based research, where Danish project developers only have to notify the agency and wait for its clearance to proceed with their respective projects [1, 23, 24]. Therefore, this simplified approval process for exclusively data-based research, which also encourages project developers to address data security and privacy concerns before seeking clearance from the data protection agency, can be an alternative option to the "consent or anonymize" approach used in the Swiss context [1, 20]. This is even more pressing given that privacy concerns are likely to be exacerbated by the implementation of precision health approaches, since their proper functioning requires that large quantities of health data are shared among key stakeholders [25].

Another interesting finding of the systematic review was that political barriers were only reported for Swiss national projects [1]. In contrast to Denmark where data processing practices are well established and institutionalized [26], Swiss project developers often need to cater for the different interests of the parties involved. In the same vein, one could find an important and illustrative example of how the entry in force of the *Bundesgesetz über das elektronische Patientendossier*[27] was delayed due to the diverging priorities and interests from the multitude of stakeholders involved (e.g. health insurers, health professionals and patient organizations to name a few) [14, 27].

Fairness of data sharing activities

Chapters 4 and 5 of this thesis have provided empirical insights on the fairness of data sharing activities for the original data collectors, and identified two levels, namely systemic and individual, where issues of fairness have to be addressed and proper incentives developed in order to promote the sharing of health data. Addressing these two levels of fair data sharing should help in sustaining the open science and data sharing movements that are being endorsed in the Swiss context [28]. These findings are of utmost importance since they also support the DORA[28] recommendations [29] and provide insights on not only how some recommendations could be undermined by certain systemic and individual factors, but also how these could be operationalized in practice under proper safeguards. These findings are also well-timed as they can inform and potentially facilitate the implementation of the abovementioned recommendations in Switzerland. Indeed, Swiss funders (e.g. SNSF[29] and SAMS[30]) have signed the declaration [30, 31] and are currently implementing its recommendations through different pilot projects (e.g. the SNSF is currently testing in medicine and biology an innovative scientific CV, called "SciCV", which requires information on and therefore recognizes the importance of various types of scientific outputs (including datasets) whilst excluding "journal-based metrics" [32]).

Systemic fairness for sharing health data

In this regard, chapter 4 provides an important contribution by underlining systemic tensions between the movement of open science and data sharing instilled by journal, governmental and funding organizations – as described above – and academic survival in the increasingly hypercompetitive Swiss academic environment. The effects of hypercompetition on research

[27] German for: Federal Law on the Electronic Patient Record
[28] Acronym for: Declaration on Research Assessment; also known as the San Francisco DORA
[29] Acronym for: Swiss National Science Foundation
[30] Acronym for: Swiss Academy of Medical Sciences

practices have been widely discussed [33-37]. The latter has encouraged some to embrace individualism and subsequently refrain from sharing their datasets to maintain their competitive advantage[31]. Therefore, there is an urgent need to review the academic reward mechanism – which is also one of the recommendations in the DORA declaration [29] – to ensure that the scientific value of datasets is explicitly recognized and that researchers engaging in data sharing activities receive due credit for their efforts and contributions made to advance science. Indeed, the SNSF received in 2019 a report on "Open Research Data", which showed that around 25% of Swiss researchers do not share their data, and one of the reasons given was the absence of recognition for data sharing activities [38].

The tension between academic survival and data sharing calls for the establishment of systemic reward mechanisms that recognize the value of data sharing activities conducted by the original data collectors. This important point could help in addressing deficiencies in current academic reward mechanisms that tend to evaluate the performance of researchers based on specific quantitative performance metrics (e.g. the number of peer-reviewed publications) for career opportunities or the allocation of grants [35], with the downside of impeding data sharing practices [39]. In this regard, the importance of accrediting the data sharing activities of original data collectors have been widely discussed in many academic disciplines [39-43]. However, there are many questions and modalities that need to be addressed before data sharing could be considered as an additional criterion in the academic evaluation process.

First, an important question arises on whether including data sharing activities as an additional criterion for academic evaluation will not end up being biased in the same way as those metrics currently being used (e.g. the number of peer-reviewed publications) [35]. The downside could be that researchers would still abuse the academic reward mechanism by sharing a high number of datasets, but whose quality and value for research or clinical/public health practice would be debatable.

Second, researchers are often reluctant to share datasets in a timely manner once their first study findings are published because they have not been able to fully exploit them. The reasons given for not sharing immediately their datasets include the resources (financial, human, technological, etc.) and efforts they have invested in the data collection and curation processes, and the fear of getting scooped by external researchers. However, it is important to comprehend that the expected secondary publications from the original data collectors are actually low in

[31] Geneviève LD, Martani A, Perneger T, Wangmo T, Elger BS. Systemic Fairness for Sharing Health Data: Perspectives From Swiss Stakeholders. Frontiers in Public Health. 2021;9.

comparison to the number of papers by third parties that would have reused their data (and cited both the paper and the dataset of the original data collectors), if the latter was shared [44].

Third, it is becoming more problematic for original data collectors to refrain from sharing datasets since funders are now pro-actively tackling this issue by requesting data management and sharing plans in research proposals as an additional prerequisite for obtaining grants. For instance, the SNSF has started requesting these data management plans in research funding applications since October 2017 [45], whilst the US National Institutes of Health (NIH) has adopted a new data management and sharing policy that will be effective as of January 2023 for all research studies funded by NIH grants [46]. Therefore, it becomes even more urgent to ensure that data sharing activities are perceived as being fair to researchers. Otherwise, high quality secondary research might be compromised by researchers providing low quality datasets that only satisfy the minimum criteria for sharing [43].

Fourth, hindering data sharing was the perceived legal uncertainty in Switzerland due to its fragmented and heterogeneous regulatory and legal practices. Specifically, two factors are believed to contribute to this perceived legal uncertainty. The first one being the broad scope of different scenarios that privacy laws need to cover, which subsequently makes them difficult to implement in specific circumstances [47]. Indeed, individual researchers in chapter 4 expressed some difficulties in ascertaining whether their data sharing activities (e.g. tools used in sharing data) were compliant with data protection laws. Secondly, the fragmented landscape of data protection regulations – cantonal and federal laws, a consequence of Swiss federalism – has complicated data sharing activities for researchers, in particular when data needed to be shared and analyzed across cantons or with a European research partner [48, 49]. Therefore, it can be argued that, under the current legal system, data sharing activities are not occurring under optimal conditions for researchers. Therefore, one recommendation that could potentially help in addressing legal uncertainty in Swiss healthcare and contribute to fairer data sharing conditions for researchers was discussed, namely in the form of adequacy models (e.g. data protection certification mechanisms) [47] dedicated for health research.

Individual notions of fair data sharing

Chapter 5 has investigated the individual notions of fair data sharing using a distributive justice lens [50]. Within the distributive justice principles, I chose the *desert-based principles* [51] to provide a theoretical framework for the analysis of data. Indeed, this theoretical framework provided three main classifications, namely *effort*, *compensation* and *contribution* to justify or hint towards data sharing practices that could be considered as fair or unfair during the

negotiation process between the original data collectors and the data recipients [51]. Three main themes emerged from chapter 5, namely (i) *effort* (subcategorized into two subthemes namely *data reciprocity and other reciprocal advantages*, and *transparency on data re-use*), (ii) *compensation* (two subthemes identified: *academic compensation* and *financial compensation*), and lastly, *contribution*. In this regard, chapter 5 provides important insights for promoting the sharing of health data by highlighting fair or unfair data sharing practices used by original data collectors, in particular when negotiating access to their datasets with external researchers.

For the first theme, it was noted that for the promotion of data sharing activities, efforts of original data collectors in collecting, managing and sharing datasets needed to be acknowledged through some reciprocity, either in the form of data reciprocity or as other reciprocal advantages to the original data collectors. The underlying reason influencing or promoting such data sharing practices could come from the Swiss hypercompetitive academic environment, whereby the original data collectors, data recipients and other external researchers are all competing for the same finite resources (e.g. career or funding opportunities), and where there is a lack of widely-implemented systemic attribution mechanisms for data sharing activities[32]. Therefore, the original data collectors might have judged that receiving data or other advantages from the data recipients is a fair practice, in particular if they want to remain competitive in academia. However, reciprocity has been discussed in data sharing frameworks, mostly as a matter of being ethical and fair towards data subjects or resources-limited researchers [52-54]. Chapter 5 thus provides an additional insight on how reciprocity could be operationalized between competitive partners in the form of data reciprocity.

In addition, normalizing data reciprocity as an incentive for sharing datasets could prevent free riding behavior of certain researchers who preyed exclusively on data from others and depleting these finite resources, whilst making few or no contribution to the data ecosystem of PM and PPH initiatives. Therefore, data reciprocity could be an interesting alternative whilst waiting for the adoption and implementation of systemic attribution mechanisms that recognize the scientific contributions made by the original data collectors for their data sharing activities, as discussed in chapter 4. Interestingly under data reciprocity, it can even be argued that academic competition would become one of the drivers of data sharing activities, and under this argument, this practice should be promoted. To illustrate the latter point, an analogy can be made with

[32] Geneviève LD, Martani A, Perneger T, Wangmo T, Elger BS. Systemic Fairness for Sharing Health Data: Perspectives From Swiss Stakeholders. Frontiers in Public Health. 2021;9.

other fields, where data reciprocity was proposed to cater for the asymmetric competition between banks and big tech companies regarding data sharing activities imposed by the "Access to Account" rule (XS2A) of the European Payment Service Directive 2 (PSD2) [55, 56]. To tackle the issue of banks being competitively disadvantaged with regard to big tech companies, which might take advantage of the unilateral flow of data from banks imposed by the XS2A rule, Di Porto and Ghidini have also argued in their article that complementing the XS2A rule with a "data reciprocity clause" would facilitate data sharing and improve competition [55]. However, unless dealing with completely anonymized datasets for which no consent is required under the HRA [21], explicit consent would have to be sought from data subjects every time data reciprocity would be required. This needs to be taken into account by both parties involved.

Another subtheme identified was a perceived claim to *transparency on data re-use*, whereby original data collectors sharing their datasets felt entitled to receive information on all aspects related to the processing, re-use and storage of their datasets by the data recipients. The latter was commonly operationalized in practice by signing contractual agreements between the parties involved [57]. With regard to common notions of transparency used in data sharing frameworks in the form of data availability statements requested to the original data collectors [57-59], our study demonstrated another important transparency dimension. This dimension deserves to be further explored in other studies with regard to the setting up of standardized templates specific to health research, which would facilitate and streamline such requests from data recipients for an easy flow of data, whilst safeguarding the interests of not only the original data collectors but also those of data subjects.

Under the *compensation* theme, another important finding from chapter 5 concerns the claim to *academic compensation*, whereby original data collectors felt entitled to receive co-authorship or collaboration opportunities from data recipients for their shared datasets. Such data sharing practices have been considered as being fair by scientists from different fields [60], and the ICMJE [61] has also recommended that collaboration or co-authorship opportunities could be offered to acknowledge the contributions made by the original data collectors. However, simply sharing datasets with data recipients do not justify co-authorship in the latter's publications. The ICJME also made it clear that four authorship criteria need to be fulfilled [62] in order for original data collectors to qualify as co-authors in subsequent publications reusing their data. However, if authorship criteria are not fulfilled, another satisfying alternative could reside in the concept of "data authorship" as proposed by Bierer, Crosas and Pierce [43].

Another subtheme identified was a claim to *financial compensation* by the original data collectors, whereby the latter felt entitled to receive some monetary compensation to recover, at least, part of the costs incurred in processing their respective datasets. Some participants further argued that such an approach would also help in ensuring the longevity of health databases, whilst maintaining adequate quality standards for shared datasets. The fairness of such financial claims to accessing data did not reach unanimity among scientists, as shown by the international survey carried out by Tenopir and colleagues, where the majority of participants were against it [60]. If ever, the claim to financial compensation becomes an established practice, it is crucial to ensure that financial interests do not guide data sharing activities, but rather favors the recovery of costs incurred by the original data collectors during their data processing activities [63]. This highlights the need to establish some standardized metrics to calculate transparently and fairly data access fees for the data recipients, as discussed in chapter 5.

The last theme identified but rarely mentioned by study participants, *contribution*, was a claim to compensation mechanisms based on the importance and quantity of contributions made by the original data collectors. An interesting anecdote conveyed by one of the study participants concerned the determination of authorship order in a manuscript based on the quantity of data contributed by each original data collector. Although in health sciences, authorship order reflects distributive justice principles [64], the position that the quantity of contributed data provides a fair and appropriate basis to assign authorship order is a reductionist view of the expected author contributions in a given manuscript. Such practice should not be promoted, as discussed in chapter 5. The main obstacles and potential solutions/fair practices to the fair sharing of health data from both a systemic and individual perspective are summarized in figure 1.

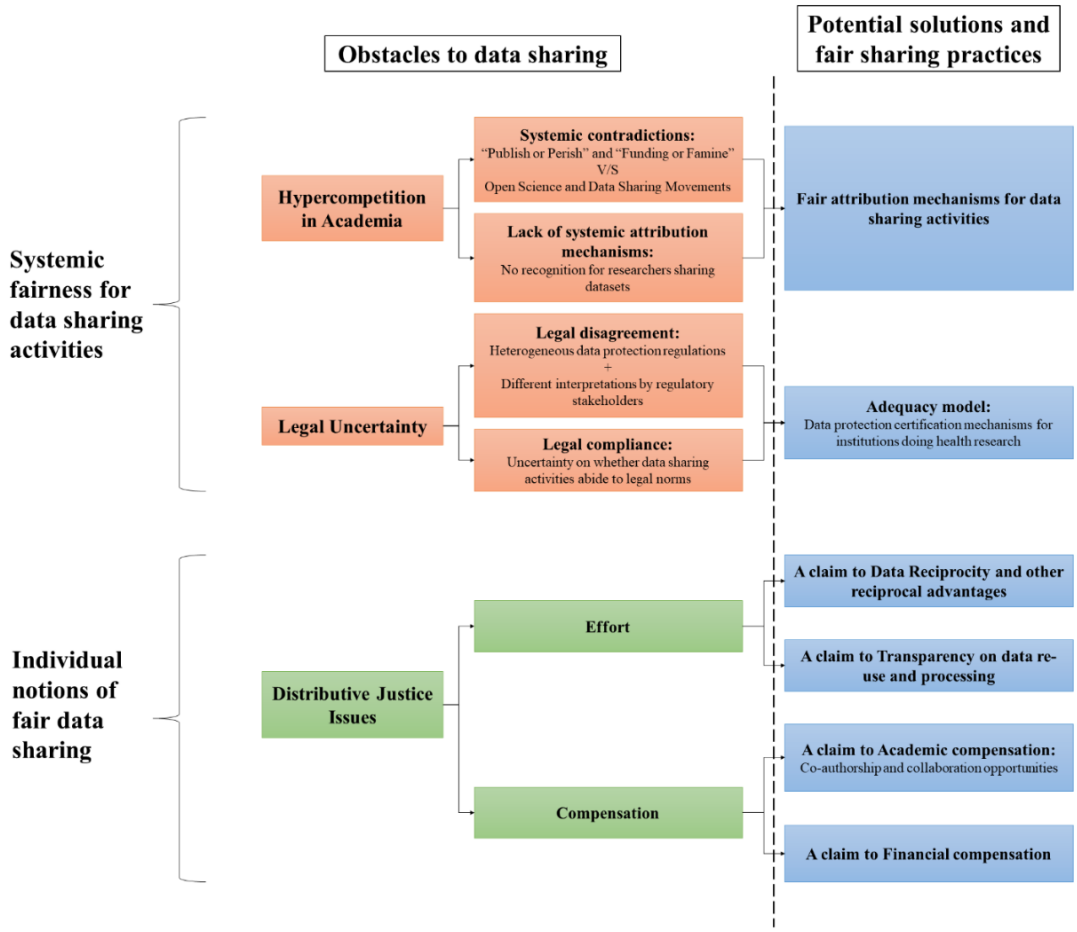


Figure 1 Summary findings of Chapters 4 and 5

With regard to PM and PPH approaches, addressing the mindset of researchers regarding data sharing activities is paramount for the latter to achieve their respective goals, requiring in particular "moving away from this closed 'selfish silo' approach to embrace a more 'open source' collaborative culture" [25]. I believe that a better understanding of the obstacles, potential solutions and identified fair data sharing practices discussed in chapters 4 and 5 could be the foundation of such transformational endeavor by recognizing and valuing the interests of the original data collectors. On this matter, further research is required.

10.3 ELSIs and Research Ethics for Digital Participatory Disease Surveillance Systems

Chapters 6 and 7 of this thesis have explored, analyzed and suggested solutions to the ELSIs arising from the involvement of digital communities of volunteer citizens who report syndromic information on participatory disease surveillance systems (online platforms or via mobile apps) of the European Influenzanet Consortium. During the course of these two chapters, these following ELSIs were addressed and analyzed.

The validity of informed consent in digital public health surveillance

As argued in the introduction section of this thesis, the traditional way of obtaining informed consent is ill-adapted to the digital environment for numerous reasons (e.g. within the framework of *Big Data*, the future potential uses and associated risks of the collected data would be unknown at the time consent is sought from participants, thereby undermining the informed nature of the consent process) [65, 66]. With respect to participatory systems, this also includes the high numbers of potential participants where it would have been practically and financially challenging for researchers to obtain valid informed consent [67]. In this regard, chapter 6 has shown that most of the country-specific Influenzanet platforms have opted for an electronic adaptation of traditional consent, known as an e-consent or electronic consent, paired with the ability of participants to opt out from the online studies [3, 68]. However, as chapter 7 has highlighted, this online form of consent also presents some vulnerabilities that necessitated some additional safeguards to protect individual interests, the trustworthiness, sustainability and integrity of these surveillance platforms. These are elaborated below and further discussed in accordance with relevant literature.

First, obtaining informed consent through electronic means presents unique challenges in comparison to more traditional forms of consent. Indeed, the interaction between study participants and researchers happens virtually, and where consent is presumed to be sensibly obtained based on some approved set of computerized actions made by the participants (e.g. by clicking through a set of buttons in the e-consent form), with the additional risk that in the best scenario the information has been understood partially, or in the worst case, not even understood or read [68, 69]. Second, an additional challenge is raised by this virtual interaction between researchers and participants in e-consent procedures: the validity of the consent procedure is undermined by the difficulties in ascertaining whether the participant has the legal capacity to provide a meaningful informed consent because of the difficulties associated in proving his or her identity [69].

Therefore, recommendations were formulated in chapters 6 and 7 in the form of information being concisely summarized as a single reader-friendly PDF file and requiring participants to digitally sign the consent form when registering on the platform and each time, new information is added. Additionally, the validity of the e-consent procedure can also be fostered through the application of ethics guidelines developed by the *British Psychological Society* (discussed in chapter 6) [70]. Such measures will motivate participants to read the informed consent information and helped in ascertaining their identity. Furthermore, the use of gamification such

as quizzes and other tools (e.g. self-explaining pictures or videos) could help participants understand more comprehensively the informed consent information [71].

Data security and protection of participants' privacy

Chapters 6 and 7 provide insights on how country-specific platforms of the Influenzanet Consortium guarantee the privacy and security of participants' data, namely through: (i) anonymization of collected datasets and their storage in a centralized and protected database [72], (ii) the transmission of postal code data to the Influenzanet platform rather than more fine-grained location data for potential cases [73], and (iii) the processing of sensor data on participants' mobile phones and transmission of results only in highly aggregated and anonymized formats to their respective Influenzanet platforms [74]. In spite of these measures taken by the Influenzanet platform, it is important to highlight that the risk of re-identification of data subjects still remains. For instance, in the event of a data leakage and with the advent of new technologies allowing the coupling and analysis of those health data collected for epidemic forecasting studies with other available databases (e.g. local data), it is possible that not only individuals could be re-identified but also that the latter and their neighborhoods could end up being stigmatized [65, 75-77]. Moreover, one common characteristic of epidemics forecasting studies is that the collected datasets are often stored for a long time and gradually enriched with additional updated information over time. Although essential in modeling disease progression, such characteristic also exposes data subjects to an increased re-identification risk [78]. Therefore, in the case of participatory disease surveillance systems and other digital public health surveillance systems, anonymization is definitely no *silver bullet* in guaranteeing the ethical processing of data gathered from study participants [65]. Furthermore, accountable and responsible sharing of anonymized datasets collected for public health surveillance and research should be promoted to guard against any re-identification risks and misuse of those datasets [4].

Distributive justice issues in the broad societal context

Distributive justice [50] demands that the benefits and burdens of participation in public health surveillance are fairly distributed between study participants and society at large, whilst taking into consideration their necessities and rights. From the analysis of the Influenzanet country-specific platforms carried out in chapters 6 and 7, it was observed that participation is open, voluntary, free and non-discriminatory for all except one of these platforms [73], which aligned with distributive justice principles. Although laudable in terms of fulfilling distributive justice claims, another aspect that could undermine the fairness of such platforms resides in the fact that different societal groups have different inclinations and access to digital technologies, a limitation known as the "digital divide" [65, 79]. In this vein, the digital divide will influence

the representation of different societal groups in the epidemics forecasting studies: e.g., in Influenzanet studies, participants were predominantly middle-aged and with a higher than average educational background [80].

This raises the question of whether digital public health surveillance is not contributing to the exacerbation of health inequities among the different societal groups, undermining distributive justice principles, with the latter two described groups (middle-aged and high education) being more prone to benefit from its interventions [65]. However, members of the Influenzanet Consortium have catered for such disparities by providing information on epidemics via other means to ensure a fairer distribution of the benefits of their interventions to those societal groups not inclined to digital technology. Examples include the dissemination of information through the use of press releases, weekly bulletins at the national level and radio broadcasting [81, 82]. Being a public good, it can be argued that the benefits of public health surveillance are likely to be shared with society at large, rather than solely benefitting participants who contributed their data to the national surveillance platforms [83].

The concept of solidarity [84] also dictates that more granular and fine-grained data on the progression of influenza epidemics, which are up-to-now only available to Influenzanet participants [81], should also be made available to non-participants. Indeed, epidemics do not differentiate between participants and non-participants, and the combined protection of both would alleviate the yearly mortality and morbidity burden associated with seasonal influenza, and confer a better protection against potential pandemics.

Capacities of Research Ethics Committees in digital public health surveillance

In chapter 7, it is discussed that new technological advances undermine the efficiency of some national regulatory frameworks to safeguard the interests of data subjects, in particular given their latency to adapt to the rapidly developing technological field. Specifically, definitions of widely-accepted and implemented safeguards to minimize privacy and data security risks, such as data *anonymization* or *encryption,* are being challenged by these new technologies (e.g. with big data analytics where re-identification of completely anonymized datasets was proved to be possible) [85]. Therefore, it is argued in chapter 7 that RECs need to cater for these regulatory gaps by engaging more actively in the design and execution phases of online disease surveillance systems, such as those of a participatory nature or those involving the use and analysis of *Big Data*. In order to fulfill their research governance for digital public health surveillance, it was further argued that RECs have to undergo interdisciplinary capacity building through training and exchange of information between the different key stakeholders

involved (e.g. data and citizen scientists, ethicists, *Big Data* researchers, to name a few). Such capacity building will contribute to ensure that RECs are able to safeguard the interests of data subjects, whilst avoiding the formation of unnecessary barriers to the use of such innovative public health interventions [4]. This particular recommendation is supported by Ienca and colleagues [86], who have also argued for expanding the technical and methodological capacities of RECs with regard to health research involving *Big Data*, given its novel ethical challenges.

Heterogeneous ethical approaches: implications for research ethics

In chapter 6, it was shown that the way research ethics are implemented throughout the country-specific platforms of the Influenzanet Consortium needed to be harmonized (e.g. three of the 11 country-platforms did not seek research ethics approval and two platforms have not applied e-consent procedures to authorize data collection from their participants). Therefore, the chapter underlined heterogeneous evaluation processes despite similar, if not identical, data processing activities [3] (also underscored in chapter 5 for certain research projects owing to some legal uncertainty). Unfortunately, these heterogeneous evaluation processes and application of research ethics could contribute to obstacles in the sharing of data across national borders, in particular if some local RECs judge the evaluation processes made by external ones as insufficient to guarantee and protect the interests of local data subjects. Therefore, such ethical approaches need to be harmonized throughout the consortium: the evaluation process by RECs, the need for electronic consent to authorize data processing, standardized informed consent information, standardized data processing and data security measures in place to protect the privacy of data subjects. Data security measures could also be standardized through an adequacy model (discussed in chapter 4), whereby each platform would have to undergo the same data protection certification mechanism to promote data sharing between them [47].

The ELSIs and their solutions for digital participatory disease surveillance systems are summarized in figure 2.

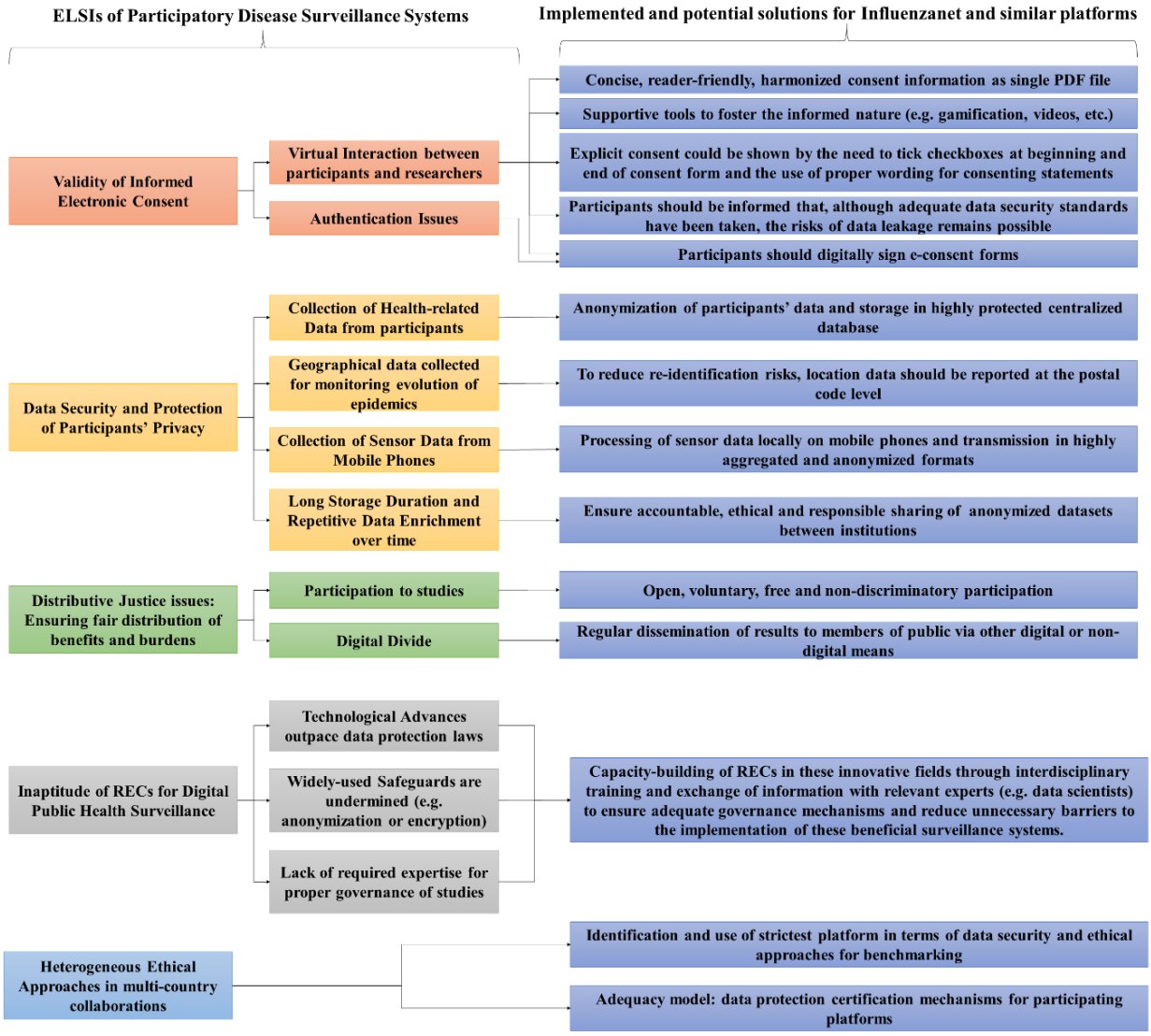


Figure 2 ELSIs and solutions for digital participatory disease surveillance systems

10.4 Racial and ethnic discrimination in precision health approaches: finding a better way forward to achieve health equity and social justice

After discussing general issues that influenced the processing of health data in previous sections, chapters 8 and 9 offer a different assessment by tackling the important and timely topic of structural racism in precision health approaches, in particular forecasting its impact on data processing activities and seeking solutions to ensure that precision health approaches abide to their social justice and health equity goals. In addition, careful consideration was given to the increasing use of AI technologies (such as machine learning, discussed in section 1.3 of the introduction) in these precision approaches to health.

Structural racism in precision public health: concerns and potential solutions

In chapter 8, it was underscored how deeply entrenched structural inequalities scourging racial and ethnic minority groups contributed to the disproportionate impact of the novel coronavirus

virus in these communities, particularly visible in countries where data on racial and ethnic origin were available to inform public health practice (e.g. the US or the UK) [87-89]. Indeed, in the US, COVID-19 age-adjusted mortality rates for some racial and ethnic minorities were at minimum three times higher than for the Caucasian population [89]. For minority groups, contributing factors to the observed increased death toll included a diminished access to quality care and COVID-19 testing, difficulties to abide to public health recommendations due to associated structural vulnerabilities (e.g. residential segregation, redlining policies, job insecurity, lower insurance coverage, poverty and multigenerational houses), and an increased prevalence of health conditions that predispose racial and ethnic minorities to an increased risk of health complications and death from COVID-19 [87, 90-94]. Many of these contributing factors – if not all – are direct or indirect consequences of structural racism.

In addition, another facet of injustice was brought in broad daylight: the implementation of color-blind policies to fight the pandemic due to the suboptimal quality and non-systematized collection of racial and ethnic data that could have informed and increased the efficiency of public health interventions for these disadvantaged and marginalized societal groups [92, 95, 96]. Similarly, in continental Europe, color-blind policies against the pandemic are likely to be implemented due to legal obstacles that prohibit the processing of racial and ethnic data [97, 98]. Therefore, the question also arises – as in the US – whether such policies are not in fact, concealing the impact of the virus on racial/ethnic minority populations and undermining the effectiveness of the public health response [99]. Addressing this data crisis is even more pressing given that PPH is being considered as an additional tool that could benefit the fight against the COVID-19 pandemic [100]. In this regard, chapter 8 discusses the two foundations of PPH approaches whilst referring to the impact of structural racism, namely their reliance on data and technology [101, 102]. Subsequently, it forecasts that PPH is likely to exacerbate health inequities between racial and ethnic groups if proper solutions are not implemented to counter the effects of structural racism.

In chapter 8, it is argued that structural racism is likely to have an impact on PPH approaches, in the form of data racism [103] and with the use of machine learning technologies (the three main classes of machine learning techniques [104, 105] are discussed in section 1.3) [106]. In addition, the role of socio-political factors on public health institutions was also addressed, in particular how the former could foster the racialization of the latter (for instance, by limiting the employment of minorities in high positions within the institutions) or hindering public health responses aimed at promoting health equity [107]. Therefore, recommendations to

address issues of data racism, algorithmic bias and socio-political factors were put forward[33]. In addition, other solutions devised for PM in chapter 9 could also work for PPH approaches. These are discussed in the next section.

Structural racism in precision medicine: concerns and potential solutions

As stated above, chapter 9 formulates some potential actions at each node to try reduce the negative effects that structural racism could have on the generation and integration of health data from racial/ethnic minorities, and on the deliverables of PM. The important contributions of this chapter were to provide a thorough analysis of how the processing of unrepresentative data from racial/ethnic minority groups induced by structural racism could lead to adverse discriminatory outcomes for the latter in PM initiatives, thereby undermining rights to social justice and health equity, and subsequently, proposed potential actions to remedy them [5].

The first node concerns the initial interaction between racial/ethnic minorities and their respective healthcare system or researchers, which leads to the generation of biased datasets [5]. It was highlighted in chapter 9 that the impact of structural racism on this initial encounter is twofold, first in the form of biased guidelines and treatment options due to the underrepresentation of minorities in their elaboration [108], and second, the often substandard quality of care offered to minority groups due to implicit provider bias [109]. Regarding the latter, the contribution of negative stereotypes [110] on access and quality of care was highlighted in the context of PM, whereby healthcare providers might unconsciously not prescribe the newly developed therapeutics to minorities and offer only substandard treatment options whilst contributing to undermining patient trust and loss of interest in engaging in clinical decision-making [111, 112]. Such detrimental aspects undermine the efficiency of PM initiatives by producing biased and non-representative datasets on racial/ethnic minority groups, who are in turn less likely to engage in PM initiatives due to the loss of trust and the perceived suboptimal treatment options [5]. Our forecasting analysis is supported by other studies conducted in the field [113].

Another study carried out by Green and colleagues [114] was one of the first studies to investigate empirically whether implicit provider bias was likely to influence treatment decisions through *race preference implicit association tests* (IAT). Importantly, the study revealed that, although healthcare providers did not exhibit any explicit racial biases, they still harbored unconsciously racial biases favoring white over black patients [114]. It is, however,

[33] More details on the four recommendations are available in chapter 4

encouraging to know that such racial/ethnic biases, although unconscious, could still be addressed and reduced via different strategies due to their malleability [115].

Therefore, our study [5] argued for the pressing need to address such implicit racial/ethnic biases and the underrepresentation of minority groups in PM initiatives through the following actions. These include prioritizing the recruitment of historically marginalized societal groups (including racial and ethnic minorities) to cater for their underrepresentation in databases destined for PM initiatives (such as in genomic studies [116]), ensuring an adequate representation of these minorities as researchers and healthcare providers [117], and providing implicit bias training for researchers and healthcare professionals through the use of IAT tests [114]. With regard to the first potential action, the importance of ensuring adequate representation of racial and ethnic minorities was highlighted in an editorial by Daniel Spratt [118], which also supports our findings to prioritize the recruitment of minority groups for PM. In addition, it is important to comprehend that increasing the representativeness of racial and ethnic minorities in datasets destined for PM initiatives is also a moral duty [119].

The second recommendation for this node is to increase the representation of racial and ethnic minorities in healthcare professions and as researchers in PM initiatives to help reduce the impact of structural racism [5]. This is supported by the fact that increasing patient-provider racial and ethnic concordance has been shown to provide additional benefits for the quality of care (e.g. better communication, more trustful relationships, and better adherence to treatment recommendations, to name a few) [120-123]. For instance, Saha and Beach [120] have explicitly connected the benefits of patient-provider racial and ethnic concordance for Black patients in relation to structural racism, but they also highlighted that the race of physician was not problematic to White patients.

The second node of the process flow in chapter 9 concerns the integration of biased data produced on racial and ethnic minority groups into the framework of PM initiatives [5]. The consequences of such biased integration could result in misuse in clinical and research practices [108]. In this section, the implications of feeding biased data on racial and ethnic minorities in AI technologies (especially for machine learning algorithms) are forecasted, in particular given the importance of AI for PM initiatives as discussed in the introduction section of this thesis, and the potential possessed by such technologies to worsen existing biases and discrimination to a much larger extent [124, 125]. To tackle the impact of structural racism at the second node, three potential actions have been put forward in chapter 9, namely (i) the setting up of *Algorithmic Impact Assessments* [126], (ii) the use of the concept, *datasheets for datasets,* for

selection of appropriate datasets [124], and lastly, the use of the toolkit developed by IBM, *AI Fairness 360*, for identifying, mitigating and correcting algorithmic biases [127]. In this regard, it is important to understand that such recommendations are not aimed at tackling only the inherent technical limitations or flaws of machine learning algorithms but rather aimed at also limiting or correcting at best the influences of social biases on the latter [128], instilled by structural racism. Indeed, as argued by Wachter, Mittelstadt and Russell [128]:

> "Adding more data will paint a more accurate and nuanced picture of the unequal world we live in for an algorithm to learn from or make decisions about, but it cannot resolve the root cause(s) of inequality; only individual, societal, or institutional change can. This is a feature of many technical fixes deployed in 'fair machine learning': they are a temporary fix for the symptoms, but not causes, of inequality in society" [128] (p. 8-9).

This statement also supports our multi-component approach to addressing issues at the three nodes of the process flow to anticipate and reduce the impact of structural racism on PM initiatives [5].

The third node of the process flow in chapter 9 concerns the influences of structural racism on deliverables of PM initiatives [5]. It was argued that structural racism might influence the research and pharmaceutical domains to be more productive towards finding new preventive or therapeutic options for diseases more prevalent in the majority racial and ethnic group, at detriment of minorities [5]. To tackle these potential influences of structural racism, two main recommendations were made at this particular node, namely (i) the creation and management of biobanks and other research data banks for and led by racial/ethnic minorities [5], and (ii) favoring community-based research [129]. Such measures could ensure that research on diseases affecting predominantly racial/ethnic minority groups are not neglected whilst safeguarding the participation and trust of the latter in PM initiatives [5].

10.5 General limitations

In this section, chapter-specific limitations are not discussed since they have already been highlighted in their respective chapters. Here, I present the limitations of the thesis itself.

First, three sources of data are analyzed in this thesis: interviews, reviews and theoretical work, all having some inherent limitations based on the methodologies used. Therefore, the overall findings of this thesis are not generalizable but remain nonetheless informative.

Second, although the thesis is framed under precision health approaches and the processing of health data, its chapters remain nonetheless loosely connected with one another (e.g. data

processing, structural racism in precision health approaches and participatory disease surveillance systems). However, this reflects the academic freedom I have benefitted during my PhD work, where I was allowed to investigate some research questions that I considered important.

Third, some interpretations and recommendations made in this thesis might have been influenced unconsciously by my cultural and racial/ethnic background, being a minority group member and a foreigner.

10.6 Implications for future research

This thesis has covered a broad scope of topics – from a practical level (barriers and facilitators to data processing) to more meta-level concerns (fairness of data processing for researchers and discrimination against racial/ethnic minorities) – in healthcare and research, two domains particularly important for precision health approaches. In this regard, some implications for future research have already been recommended in previous sections of chapter 10. In addition to these research recommendations, one aspect that is outside my expertise and therefore not covered extensively in this thesis concerns the feasibility (e.g. whether data protection certification mechanisms could indeed be adapted to health research, given that they were originally intended for trade[34]) and the financial sustainability of the proposed solutions and recommendations to solve or mitigate the identified health data processing problematics. Therefore, further research is required at these levels. In the same vein, it would be interesting to investigate how public-private partnerships, increasingly considered for PM initiatives [130], would influence the fair sharing of health data. Indeed, academic science and industrial science may have different positions regarding the sharing of data due to their different goals and reward mechanisms (e.g. academic recognition versus financial profits).

On a different note, further research is also required to demonstrate empirically the presence and impact of structural racism on precision health approaches. In order to fulfill this objective for the Swiss context, I will be investigating, as of June 2021, the presence and impact of structural racism in Swiss precision oncology initiatives through a focus group study with racial/ethnic minority cancer patients, patient representatives, researchers and oncologists involved in precision oncology programs. However, it would be interesting to also investigate whether structural racism could have an impact in PM initiatives other than those from the

[34] Asia-Pacific Economic Cooperation (2019): "What is the Cross-Border Privacy Rules System?". URL: https://www.apec.org/About-Us/About-APEC/Fact-Sheets/What-is-the-Cross-Border-Privacy-Rules-System. [Accessed date: November 5, 2020]

Western world (e.g. PM initiatives in India, China and African countries), especially in those countries which were deeply impacted by past colonial powers.

10.7 Conclusions

This thesis has achieved its research objectives, that is, (i) it has provided insights on the barriers and facilitators that are influencing the processing of health data in the Swiss context; (ii) it has investigated and analyzed the different ELSIs arising from participatory disease surveillance systems and suggested ways to improve ethical approaches by providing a robust ethical framework; and lastly, (iii) it has promoted discussion on the impact of structural racism on precision health approaches, identified vulnerabilities in their data ecosystems, and thereafter, proposed solutions and recommendations to help such initiatives abide to their health equity and social justice endeavors. Regarding the processing of health data in the Swiss context, the thesis has demonstrated that Swiss research projects have to face different types of obstacles in their data processing activities. Additionally, it reveals that specific characteristics of the Swiss context have influenced the design of projects, the latter being more prone to the prospective collection of health data. This implies that Swiss projects are likely to encounter specific obstacles, which will necessitate the development of additional solutions in comparison to those carried out in Denmark. This observation highlights that it is crucial to understand comprehensively the country-specific contextual particularities within which research projects are being developed. In addition, this thesis provides important insights on the existence of systemic contradictions to the sharing of health data, which highlights the need to improve the fairness of academic reward mechanisms. Moreover, fair data sharing practices should be promoted in particular if research datasets cannot be made available on open repositories. Such measures should help in ensuring that these datasets are made available and can inform PPH and PM approaches. Lastly, this thesis discusses the important and timely topic of structural racism in precision health approaches. Its recommendations might help in preventing the reiteration of the *status quo* observed in healthcare and public health for these emerging fields.

10.8 References

1. Geneviève LD, Martani A, Mallet MC, Wangmo T, Elger BS. Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: A systematic review. PLoS one. 2019;14(12):e0226015.

2. Hansen J, Wilson P, Verhoeven E, Kroneman M, Kirwan M, Verheij R, et al. Assessment of the EU Member States' rules on health data in the light of GDPR. Brussels: DG Health and Food Safety; 2021.

3. Geneviève LD, Wangmo T, Dietrich D, Woolley-Meza O, Flahault A, Elger BS. Research Ethics in the European Influenzanet Consortium: Scoping Review. JMIR Public Health Surveill. 2018;4(4):e67.
4. Geneviève LD, Martani A, Wangmo T, Paolotti D, Koppeschaar C, Kjelsø C, et al. Participatory Disease Surveillance Systems: Ethical Framework. J Med Internet Res. 2019;21(5):e12273.
5. Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. BMC Med Ethics. 2020;21(1):17.
6. Meier-Abt P, Lawrence A, Vayena E, Schwede T. The Swiss approach to precision medicine. Swiss Med Wkly. 2018.
7. Basin D. Analysis on Different Usages of the Swiss AHV Number (2017) [accessed 2017]. Available from: https://www.bj.admin.ch/bj/de/home/publiservice/publikationen/externe/2017-09-27.html.
8. De Pietro C, Camenzind P, Sturny I, Crivelli L, Edwards-Garavoglia S, Spranger A, et al. Switzerland: health system review. Copenhagen: World Health Organization. Regional Office for Europe; 2015 2015.
9. Wallace LS. A View Of Health Care Around The World. Ann Fam Med. 2013;11(1):84.
10. Crivelli L, Salari P. The inequity of the Swiss health care system financing from a federal state perspective. Int J Equity Health. 2014;13(1):17.
11. Federal Office of Public Health (FOPH). Directories of approved health insurers and reinsurers (2018) [cited 2019]. Available from: https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung/krankenversicherung-versicherer-aufsicht/verzeichnissekrankenundrueckversicherer.html.
12. OECD. Strengthening health information infrastructure for health care quality governance: good practices, new opportunities and data privacy protection challenges: OECD Publishing; 2013.
13. Braendle T, Colombier C. Budgetary targets as cost-containment measure in the Swiss healthcare system? Lessons from abroad. Health Policy. 2020;124(6):605-14.
14. De Pietro C, Francetic I. E-health in Switzerland: The laborious adoption of the federal law on electronic health records (EHR) and health information exchange (HIE) networks. Health Policy. 2018;122(2):69-74.
15. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort Studies: Prospective versus Retrospective. Nephron Clin Pract. 2009;113(3):c214-c7.

16. Institute of Medicine (US). 3, Barriers to Data Sharing. Sharing Clinical Research Data: Workshop Summary. Washington (DC): National Academies Press (US); 2013.
17. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. Eur J Hum Genet. 2018;26(2):149-56.
18. Siu LL, Lawler M, Haussler D, Knoppers BM, Lewin J, Vis DJ, et al. Facilitating a culture of responsible and effective sharing of cancer genome data. Nat Med. 2016;22(5):464-71.
19. Townend D. Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? Hum Genet. 2018;137(8):657-64.
20. Mostert M, Bredenoord AL, Biesaart MC, van Delden JJ. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet. 2016;24(7):956-60.
21. Schweizerische Eidgenossenschaft. Loi fédérale relative à la recherche sur l'être humain: L'Assemblée fédérale de la Confédération suisse; 2021 [accessed 6.02.2021]. Available from: https://www.fedlex.admin.ch/eli/cc/2013/617/fr.
22. Intersoft Consulting. Recital 26: Not Applicable to Anonymous Data*: Intersoft Consulting; (n.d.) [accessed 6.02.2021]. Available from: https://gdpr-info.eu/recitals/no-26/.
23. Copenhagen Healthtech Cluster. Better use of health data 2018 [cited 2019]. Available from: https://www.cphhealthtech.com/~/media/chc/downloads/pdf/data%20redder%20liv%20cases/better-use-ofhealth-data-report-english-version.ashx?la=en.
24. Statistik Denmark. Step-by-step practices for national register-based research projects at Statistics Denmark 2018 [cited 2019]. Available from: https://www.dst.dk/en/TilSalg/Forskningsservice.
25. Lawler M, Maughan T. From Rosalind Franklin to Barack Obama: Data Sharing Challenges and Solutions in Genomics and Personalised Medicine. The New Bioethics. 2017;23(1):64-73.
26. Olejaz M, Juul Nielsen A, Rujkjobing A, Okkels Birk H, Krasnik A, Hernández-Quevedo C. Denmark: health system review. Copenhagen, Denmark: World Health Organization, on behalf of the European Observatory on Health Systems and Policies; 2012.
27. Schweizerische Eidgenossenschaft. Bundesgesetz über das elektronische Patientendossier (EPDG): Schweizerische Eidgenossenschaft,; 2020 [9.03.2021]. Available from: https://www.fedlex.admin.ch/eli/cc/2017/203/de.

28. SNSF. Open Science; 2021 [accessed 12.03.2021]. Available from: http://www.snf.ch/fr/pointrecherche/dossiers/open-science/Pages/default.aspx.
29. DORA. San Francisco Declaration on Research Assessment n.d. [accessed 12.03.2021]. Available from: https://sfdora.org/read/.
30. SNSF. DORA declaration 2021 [accessed 12.03.2021]. Available from: http://www.snf.ch/en/theSNSF/research-policies/dora-declaration/Pages/default.aspx.
31. ASSM. «Open Access»: pour un accès libre aux résultats de la recherche scientifique. 2014. Contract No.: 1.
32. SNSF. http://www.snf.ch/en/researchinFocus/newsroom/Pages/news-200131-scicv-snsf-tests-new-cv-format-in-biology-and-medicine.aspx: Swiss National Science Foundation,; 2020 [accessed 10.03.2021]. Available from: http://www.snf.ch/en/researchinFocus/newsroom/Pages/news-200131-scicv-snsf-tests-new-cv-format-in-biology-and-medicine.aspx.
33. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLoS one. 2015;10(8):e0134826.
34. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data Sharing by Scientists: Practices and Perceptions. PLoS One. 2011;6(6):e21101.
35. Edwards MA, Roy S. Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. Environ Eng Sci. 2017;34(1):51-61.
36. Waaijer CJF, Teelken C, Wouters PF, van der Weijden ICM. Competition in Science: Links Between Publication Pressure, Grant Pressure and the Academic Job Market. High Educ Policy. 2018;31(2):225-43.
37. Tijdink JK, Schipper K, Bouter LM, Maclaine Pont P, de Jonge J, Smulders YM. How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers. BMJ Open. 2016;6(2):e008681.
38. von der Heyde M. Open Research Data: Landscape and cost analysis of data repositories currently used by the Swiss research community, and requirements for the future. Weimar, Germany: [Report to the SNSF]; 2019.
39. Pisani E, Whitworth J, Zaba B, Abou-Zahr C. Time for fair trade in research data. Lancet. 2010;375(9716):703-5.
40. Pierce HH, Dev A, Statham E, Bierer BE. Credit data generators for data reuse. Nature. 2019;570(7759):30-2.

41. Thorisson GA. Accreditation and attribution in data sharing. Nat Biotechnol. 2009;27(11):984-5.
42. Costello MJ. Motivating Online Publication of Data. BioScience. 2009;59(5):418-27.
43. Bierer BE, Crosas M, Pierce HH. Data Authorship as an Incentive to Data Sharing. N Engl J Med. 2017;376(17):1684-7.
44. Piwowar HA, Vision TJ. Data reuse and the open data citation advantage. PeerJ. 2013;1:e175.
45. SNSF. Open Research Data: Swiss National Science Foundation; 2021 [accessed 11.03 2021]. Available from: http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx.
46. NIH. Final NIH Policy for Data Management and Sharing: National Institutes of Health, Office of Extramural Research; n.d. [accessed 11.03 2021]. Available from: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.
47. Phillips M. International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR). Hum Genet. 2018;137(8):575-82.
48. Martani A, Egli P, Widmer M, Elger B. Data protection and biomedical research in Switzerland: setting the record straight. Swiss Med Wkly. 2020;150:w20332.
49. Eiss R. Confusion over Europe's data-protection law is stalling scientific progress. Nature. 2020;584(7822):498.
50. Kaufman A. Distributive Justice, Theories of. In: Chadwick R, editor. Encyclopedia of Applied Ethics (Second Edition). San Diego: Academic Press; 2012. p. 842-50.
51. Lamont J, Favor C. Distributive Justice: Stanford Encyclopedia of Philosophy; 2017 [accessed 22.02.2021]. Available from: https://plato.stanford.edu/entries/justice-distributive/#Strict.
52. Knoppers BM. Framework for responsible sharing of genomic and health-related data. Hugo J. 2014;8(1):3-.
53. Mascalzoni D, Dove ES, Rubinstein Y, Dawkins HJS, Kole A, McCormack P, et al. International Charter of principles for sharing bio-specimens and data. Eur J Hum Genet. 2015;23(6):721-8.
54. Merson L, Phong TV, Nhan LNT, Dung NT, Ngan TTD, Kinh NV, et al. Trust, Respect, and Reciprocity:Informing Culturally Appropriate Data-Sharing Practice in Vietnam. J Empir Res Hum Res Ethics. 2015;10(3):251-63.

55. Di Porto F, Ghidini G. “I Access Your Data, You Access Mine”: Requiring Data Reciprocity in Payment Services. IIC - International Review of Intellectual Property and Competition Law. 2020;51(3):307-29.
56. The European Parliament and of the Council. Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC: EUR-Lex; 2015 [accessed 16.03.2021]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015L2366.
57. Shabani M, Obasa M. Transparency and objectivity in governance of clinical trials data sharing: Current practices and approaches. Clinical Trials. 2019;16(5):547-51.
58. Hopkins AM, Rowland A, Sorich MJ. Data sharing from pharmaceutical industry sponsored clinical studies: audit of data availability. BMC Med. 2018;16(1):165.
59. Boué S, Byrne M, Hayes AW, Hoeng J, Peitsch MC. Embracing Transparency Through Data Sharing. Int J Toxicol. 2018;37(6):466-71.
60. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PloS one. 2015;10(8):e0134826-e.
61. ICJME Editors. Clinical Trials 2021 [accessed 22.1.2021]. Available from: http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html.
62. International Committee of Medical Journal Editors. Defining the Role of Authors and Contributors 2021 [accessed 25.02.2021]. Available from: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html.
63. Cole CL, Sengupta S, Rossetti S, Vawdrey DK, Halaas M, Maddox TM, et al. Ten principles for data sharing and commercialization. J Am Med Inform Assoc. 2021; 28(3):646-649
64. Helgesson G, Eriksson S. Authorship order. Learn Publ. 2019;32(2):106-12.
65. Thiel MJ, Parkin J, Dratwa J, Halila R, Palazzani L, A. G. The ethical implications of new health technologies and citizen participation: Publications Office of the European Union; 2016 [accessed 1.10.2017]. Available from: https://op.europa.eu/en/publication-detail/-/publication/e86c21fa-ef2f-11e5-8529-01aa75ed71a1.

66. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big Data for Infectious Disease Surveillance and Modeling. The Journal of Infectious Diseases. 2016;214(suppl_4):S375-S9.
67. Elger BS, Caplan AL. Consent and anonymization in research involving biobanks. EMBO reports. 2006;7(7):661-6.
68. Coiera E, Clarke R. e-Consent: The design and implementation of consumer consent mechanisms in an electronic environment. J Am Med Inform Assoc. 2004;11(2):129-40.
69. Grady C, Cummings SR, Rowbotham MC, McConnell MV, Ashley EA, Kang G. Informed Consent. N Engl J Med. 2017;376(9):856-67.
70. Hewson C, Buchanan T, Brown I, Coulson N, Hagger-Johnson G, Joinson A, et al. Ethics Guidelines for Internet-mediated Research Leicester: British Psychological Society; 2017 [accessed 1.10.2018]. Available from: https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research%20%282017%29.pdf.
71. Simon CM, Klein DW, Schartz HA. Traditional and Electronic Informed Consent for Biobanking: A Survey of U.S. Biobanks. Biopreservation and Biobanking. 2014;12(6):423-9.
72. Paolotti D, Carnahan A, Colizza V, Eames K, Edmunds J, Gomes G, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. Clin Microbiol Infect. 2014;20(1):17-21.
73. Koppeschaar CE, Colizza V, Guerrisi C, Turbelin C, Duggan J, Edmunds WJ, et al. Influenzanet: Citizens Among 10 Countries Collaborating to Monitor Influenza in Europe. JMIR Public Health Surveill. 2017;3(3):e66.
74. Hirsch M, Woolley-Meza O, Paolotti D, Flahault A, Lukowicz P. grippeNET App: Enhancing Participatory Influenza Monitoring Through Mobile Phone Sensors. Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers; Singapore, Singapore: Association for Computing Machinery; 2018. p. 833–41.
75. Ölvingson C, Hallberg J, Timpka T, Lindqvist K. Ethical issues in public health informatics: implications for system design when sharing geographic information. J Biomed Inform. 2002;35(3):178-85.
76. Vicente CR, Freni D, Bettini C, Jensen CS. Location-Related Privacy in Geo-Social Networks. IEEE Internet Computing. 2011;15(3):20-7.

77. Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for responsible sharing of clinical trial data. N Engl J Med. 2013;369(17):1651-8.
78. Metcalf J, Crawford K. Where are human subjects in Big Data research? The emerging ethics divide. Big Data & Society. 2016;3(1):2053951716650211.
79. van Dijk J, Hacker K. The Digital Divide as a Complex and Dynamic Phenomenon. Inf Soc. 2003;19(4):315-26.
80. Cantarelli P, Debin M, Turbelin C, Poletto C, Blanchon T, Falchi A, et al. The representativeness of a European multi-center network for influenza-like-illness participatory surveillance. BMC Public Health. 2014;14:984.
81. Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, et al. Participatory Syndromic Surveillance of Influenza in Europe. J Infect Dis. 2016;214(suppl_4):S386-s92.
82. Perrotta D, Bella A, Rizzo C, Paolotti D. Participatory Online Surveillance as a Supplementary Tool to Sentinel Doctors for Influenza-Like Illness Surveillance in Italy. PLoS One. 2017;12(1):e0169801.
83. Schaefer GO, Emanuel EJ, Wertheimer A. The obligation to participate in biomedical research. JAMA. 2009;302(1):67-72.
84. Dawson A, Jennings B. The Place of Solidarity in Public Health Ethics. Public Health Rev. 2012;34(1):4.
85. Tene O, Polonetsky J. Big data for all: Privacy and user control in the age of analytics. Nw J Tech & Intell Prop. 2012;11:xxvii.
86. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. PLoS One. 2018;13(10):e0204937.
87. Yearby R, Mohapatra S. Law, structural racism, and the COVID-19 pandemic. J Law Biosci. 2020;7(1).
88. Aldridge R, Lewer D, Katikireddi S, Mathur R, Pathak N, Burns R, et al. Black, Asian and Minority Ethnic groups in England are at increased risk of death from COVID-19: indirect standardisation of NHS mortality data [version 2; peer review: 3 approved]. Wellcome Open Research. 2020;5(88).
89. APM Research Lab. The color of coronavirus: COVID-19 deaths by race and ethnicity in the U.S.: APM Research Lab; 2020 [accessed 2.12.2020]. Available from: https://www.apmresearchlab.org/covid/deaths-by-race.

90. Yearby R. Structural Racism and Health Disparities: Reconfiguring the Social Determinants of Health Framework to Include the Root Cause. J Law Med Ethics. 2020;48(3):518-26.
91. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. Lancet. 2017;389(10077):1453-63.
92. Krieger N. ENOUGH: COVID-19, Structural Racism, Police Brutality, Plutocracy, Climate Change—and Time for Health Justice, Democratic Governance, and an Equitable, Sustainable Future. Am J Public Health. 2020;110(11):1620-3.
93. Martin CA, Jenkins DR, Minhas JS, Gray LJ, Tang J, Williams C, et al. Socio-demographic heterogeneity in the prevalence of COVID-19 during lockdown is associated with ethnicity and household size: Results from an observational cohort study. EClinicalMedicine. 2020;25:100466.
94. Richardson J, Mitchell B, Meier H, Lynch E, Edlebi J. The Lasting Impact of Historic "Redlining" on Neighborhood Health: Higher Prevalence of COVID-19 Risk Factors. 2020. [accessed 2.12.2020]. Available from: https://ncrc.org/holc-health/
95. Dowling MK, Kelly RL. Policy Solutions for Reversing the Color-blind Public Health Response to COVID-19 in the US. JAMA. 2020;324(3):229-30.
96. Gravlee CC. Systemic racism, chronic health inequities, and COVID-19: A syndemic in the making? Am J Hum Biol. 2020;32(5):e23482.
97. Möschel M. Race in mainland European legal analysis: towards a European critical race theory. Ethn Racial Stud. 2011;34(10):1648-64.
98. Farkas L, Directorate-General for Justice and Consumers. Analysis and comparative review of equality data collection practices in the European Union: Data collection in the field of ethnicity. Brussels: European Commission; 2017.
99. Simon P. The failure of the importation of ethno-racial statistics in Europe: debates and controversies. Ethn Racial Stud. 2017;40(13):2326-32.
100. Rasmussen SA, Khoury MJ, del Rio C. Precision Public Health as a Key Tool in the COVID-19 Response. JAMA. 2020;324(10):933-4.
101. Weeramanthri TS, Dawkins HJS, Baynam G, Bellgard M, Gudes O, Semmens JB. Editorial: Precision Public Health. Front Public Health. 2018;6(121).
102. Horton R. Offline: in defence of precision public health. Lancet. 2018;392(10157):1504.
103. Chander S. Data racism: a new frontier: European Network Against Racism; 2020 [Available from: https://www.enar-eu.org/Data-racism-a-new-frontier.

104. AIHLEG. A definition of AI: Main capabilities and scientific disciplines. European Commission; 2019. Available from: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
105. Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. Proteomics. 2016;16(5):741-58.
106. Khoury MJ, Armstrong GL, Bunnell RE, Cyril J, Iademarco MF. The intersection of genomics and big data with public health: Opportunities for precision public health. PLoS Med. 2020;17(10):e1003373.
107. Griffith DM, Childs EL, Eng E, Jeffries V. Racism in organizations: The case of a county public health department. J Community Psychol. 2007;35(3):287-302.
108. Ferryman K, M P. Fairness in Precision Medicine. Data & Society; 2018.
109. Hall WJ, Chapman MV, Lee KM, Merino YM, Thomas TW, Payne BK, et al. Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. Am J Public Health. 2015;105(12):e60-e76.
110. Penner LA, Blair IV, Albrecht TL, Dovidio JF. Reducing Racial Health Care Disparities: A Social Psychological Analysis. Policy Insights Behav Brain Sci. 2014;1(1):204-12.
111. Penner LA, Dovidio JF, West TV, Gaertner SL, Albrecht TL, Dailey RK, et al. Aversive racism and medical interactions with Black patients: A field study. J Exp Soc Psychol. 2010;46(2):436-40.
112. Williams DR, Wyatt R. Racial Bias in Health Care and Health: Challenges and Opportunities. JAMA. 2015;314(6):555-6.
113. FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. BMC Med Ethics. 2017;18(1):19-.
114. Green AR, Carney DR, Pallin DJ, Ngo LH, Raymond KL, Iezzoni LI, et al. Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients. J Gen Intern Med. 2007;22(9):1231-8.
115. Blair IV. The Malleability of Automatic Stereotypes and Prejudice. Pers Soc Psychol Rev. 2002;6(3):242-61.
116. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538(7624):161-4.
117. Cohan D. Racist Like Me — A Call to Self-Reflection and Action for White Physicians. N Engl J Med. 2019;380(9):805-7.

118. Spratt DE. Are we inadvertently widening the disparity gap in pursuit of precision oncology? Br J Cancer. 2018;119(7):783-4.
119. Petersen KE, Prows CA, Martin LJ, Maglo KN. Personalized Medicine, Availability, and Group Disparity: An Inquiry into How Physicians Perceive and Rate the Elements and Barriers of Personalized Medicine. Public Health Genomics. 2014;17(4):209-20.
120. Saha S, Beach MC. Impact of Physician Race on Patient Decision-Making and Ratings of Physicians: a Randomized Experiment Using Video Vignettes. J Gen Intern Med. 2020;35(4):1084-91.
121. Cooper LA, Roter DL, Johnson RL, Ford DE, Steinwachs DM, Powe NR. Patient-centered communication, ratings of care, and concordance of patient and physician race. Ann Intern Med. 2003;139(11):907-15.
122. Traylor AH, Schmittdiel JA, Uratsu CS, Mangione CM, Subramanian U. Adherence to cardiovascular disease medications: does patient-provider race/ethnicity and language concordance matter? J Gen Intern Med. 2010;25(11):1172-7.
123. Penner LA, Dovidio JF, Gonzalez R, Albrecht TL, Chapman R, Foster T, et al. The Effects of Oncologist Implicit Racial Bias in Racially Discordant Oncology Interactions. J Clin Oncol. 2016;34(24):2874-80.
124. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé III H, et al. Datasheets for datasets. arXiv preprint arXiv:180309010. 2018.
125. Žliobaitė I. Measuring discrimination in algorithmic decision making. Data Mining and Knowledge Discovery. 2017;31(4):1060-89.
126. Reisman D, Schultz J, Crawford K, M. W. Algorithmic impact assessments: a practical framework for public agency accountability: AI NOW Institute; 2018 [accessed 05.06.2019]. Available from: https://ainowinstitute.org/aiareport2018.pdf.
127. Bellamy RK, Dey K, Hind M, Hoffman SC, Houde S, Kannan K, et al. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM J Res Dev. 2019;63(4/5):4: 1-4: 15.
128. Wachter S, Mittelstadt B, Russell C. Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law. W Va Law Rev. 2021.
129. Kraft SA, Cho MK, Gillespie K, Halley M, Varsava N, Ormond KE, et al. Beyond Consent: Building Trusting Relationships With Diverse Populations in Precision Medicine Research. Am J Bioeth. 2018;18(4):3-20.

130. NIH invests in next iteration of public-private partnership to advance precision medicine research for Alzheimer's disease [press release]. National Institutes of Health, March 2, 2021 2021.

Appendix

Interview Guide – SMAASH NRP-74 project

Let me begin by asking you some questions on your professional/research domain.

Q1. Could you please walk me through your professional/research activity in relation to health data?

Q2. Could you tell me more on your most recent project which currently involves the collection and sharing of health data?

Q3. Could you tell me more about these data collections you are using? Which institution provided them?
What additional types of data are being used?

Can you please describe how you acquired such data?

Q4. What is your opinion on data sharing for your project?

Q5. Have you experienced barriers towards the acquisition and/or sharing of such data?
What were these barriers and how were they addressed?

Q6. If using multiple data collections (referring to Q3):

Concerning the databases/registries you are currently working with, how did you manage to link these data sources to your data?

Q7. Have you experienced OR are you anticipating barriers towards the analysis of those data for your project? Could you elaborate more on these challenges? How were these addressed?

I would like now to switch gears and move towards legal and ethical considerations concerning data collection and data sharing and would love to learn your perspectives on those:

Q8. Do you consider informed consent for data collection/sharing? What is your strategy for obtaining it or justification for not obtaining it?

Q9. For your project, did you ever experience any legal/regulatory challenges? What were these challenges?

Did you abide to any existing national/international regulatory and ethical guidelines pertaining to your professional/research activities? If yes, which ones? How did they influence your project?

Do you see any room for improvement and if yes, what exactly?

Q10. We usually hear that institutions as well as individual researchers are not keen in sharing health data. What is your opinion on this?

Q11. In the context of your project, do you feel comfortable sharing the data you collected directly or after the first analyses, and can you explain why?

In your opinion, under what conditions would third parties be allowed to use your data?

If you agree that there are conditions under which third parties can gain access, who should these third parties be?

Q12. In light of our interesting discussion, do you have any specific recommendations you would like to make which will help to improve the health data situation in Switzerland?

Q13. Do you have any question or comment that you would like to add before we end our discussion?

It has been a pleasure knowing more about you and your research/professional activities in regard to health data. We thank you for your participation and time.

Curriculum Vitae – Lester Darryl Geneviève

PERSONAL INFORMATION

Date of birth:	May 14, 1987
Family Status:	Married
Work Address:	Institute for Biomedical Ethics, Bernoullistrasse 28 (Room 202), University of Basel, 4056-Basel
Email:	lester.genevieve@unibas.ch
Webpage:	https://ibmb.unibas.ch/en/persons/lester-darryl-genevieve/
Phone:	+41 61 207 1782
OrcID:	0000-0001-5834-2220
Resident Permit	B

EDUCATION

10/2017 - now	**Ph.D. in Biomedical Ethics**, University of Basel. Dissertation title: *Precision Health Approaches: Ethical considerations for health data processing*. Supervisors: Prof. Dr. Bernice Simone Elger (University of Basel and University of Geneva), PD Dr. Tenzin Wangmo (University of Basel); Advisor: Prof. Dr. Thomas Perneger (University of Geneva).
09/2015-06/2017	**Master degree in Global Health**, University of Geneva. Specialization in Epidemiology and Biostatistics at the Swiss Tropical and Public Health Institute (SwissTPH). Thesis: *Participatory Approaches and Open Data on Venomous Snakes: A Neglected Opportunity in the Global Snakebite Crisis?* Supervisors: Dr. Rafael Ruiz De Castaneda, Dr. Isabelle Bolon (University of Geneva)
08/2012-06/2014	**Master degree in Medicine**, University of Bordeaux
08/2007-09/2014	**Medicinae Baccalaureus, Baccalaureus Chirurgiae (MBChB)**, University of Mauritius and University of Bordeaux
08/2007 – 06/2010	**Bachelor of Science with honors in Medical Science**, University of Mauritius

Certification and continuing medical education

04/2017	Certificate in *Global Health at the Human-Animal-Ecosystem Interface*, University of Geneva Certificate in *Vaccine introduction and points for consideration*, University of Geneva
08/2015 - 09/2015	Certificate in *Challenges of Global Health*, Duke University, USA
06/2014	Certificate of accreditation in *Anaplastic Gliomas*, Yale University, USA Certificate of accreditation in *limited resource setting management of outpatient diabetes*, Harvard University, USA
03/2014 - 04/2014	Certificate of accreditation in *Alcohol, Other Drugs and Health: Current evidence,* Boston University, School of Medicine, USA
06/2011	Certificate in *Surgical Science*, St Helens and Knowsley Teaching Hospitals, NHS, UK

EMPLOYMENT AND INTERNSHIP HISTORY

2020 – now	Synthesis team member, NRP-74, Swiss National Science Foundation
2017 – now	Research assistant at University of Basel, Institute for Biomedical Ethics
2016-2017	Research assistant at University of Geneva, Institute of Global Health
2015	Pre-registration House Officer, Ministry of Health and Quality of Life, Mauritius
2013-2014	Medical Extern, University Hospitals of Bordeaux, France
2010-2011	Research assistant at Ministry of Health and Quality of Life, Mauritius

INSTITUTIONAL RESPONSIBILITIES

10/2017 – now **Institute for Biomedical Ethics, University of Basel:**

(1) Organization activities for the Contemporary Debates Seminar Series in Bioethics on the topic of "Ethics at the Edge of Medicine" (Course no.: 56848-01).
(2) Organization and teaching activities for the Contemporary Debates Seminar Series on the topic of "Smarter Healthcare: the ethics of health system efficiency" (Course no.: 5186).
(3) Research assistant in the NRP-74 project titled "Advancing SMart solutions and eliminating barriers for the Acquisition, Analysis, and Sharing of Health data in Switzerland [SMAASH]". The project aims to identify factors that influence the harmonization of health data and its collection, sharing and/or linkage to provide evidence-based recommendations for policy-making. Therefore, the project aims to promote the collection and sharing of health data between stakeholders in an ethically and legally sound manner to tackle the inefficiencies of the Swiss healthcare system.
(4) Supported several grant applications for the NRP-77 on digital transformation.

02-2017 - 05/2017 **Institute of Global Health, University of Geneva:**

(1) Research assistant in a project on participatory approaches and open data on venomous snakes of public health importance from the World Health Organization list
(2) Research assistant in a project on food markets and their dual dimension in terms of food safety and food security. The project was presented at the United Nations Conference on Habitat III in Quito, Ecuador
(3) Research and development of a crowdsourcing mobile application prototype, which was tested in the Chinese cities of Beijing and Shenzhen, in collaboration with Tsinghua University

10/2010 - 03/2011 **Ministry of Health and Quality of Life, Mauritius:**

(1) Helped in the data management of the national cancer registry
(2) Participation to the conference of the International Association of Cancer Registries (IACR).

TEACHING ACTIVITIES

Spring 2020 Institute for Biomedical Ethics: Organization of a seminar titled "Contemporary Debates in Bioethics: Ethics at the Edge of Medicine" for students of the University of Basel.

Fall 2018 Institute for Biomedical Ethics: Organization of a seminar titled "Contemporary Debates in Bioethics: Smarter Healthcare, the Ethics of Health Systems Efficiency" for students of the University of Basel.

PROFESSIONAL MEMBERSHIPS

04/2019 – now European Public Health Association (EUPHA), Member
08/2018 – now Public Health Schweiz, Member

PRIZES AND AWARDS

2020 Research grant, Käthe Zingg Schwichtenberg Fund, Swiss Academy of Medical Sciences (CHF 47,170; grant number: KZS 08/20) for the project proposal titled "Structural Racism in Precision Oncology (STRIPE): An Exploratory Study".
2020 The University of Basel Travel Fund Grant (CHF 808.00)

2016	Best Young Researcher Award, Jet d'Or de Genève, Geneva Health Forum
2016	Open Seventeen Program Call, GOVLAB (New York University) and Citizen Cyberlab (University of Geneva)

PERSONAL SKILLS

Technical	Statistical software (STATA, Epi Info), geographic information systems (QGIS), qualitative analysis software (MAXQDA)
Language	French, English (fluent, IELTS score of 8) and German (beginner)

PRESENTATION AND TALKS

2020	PEMED 2020 Personalized and Precision Medicine International Conference, poster presentation on "Structural Racism in Precision Medicine: All patients are equal but some are more equal than others", Munich, Germany (February 19-21)
2019	Swiss Society of Bioethics, selected speaker presenting on "Advancing SMart solutions and eliminating barriers for the Acquisition, Analysis, and Sharing of Health data in Switzerland [SMAASH]: Defining Priority Areas", Bigorio, Switzerland (November 28-30)
	SNSF EHCL Programme workshop, project presentation, Basel, Switzerland (March 18)
2018	Geneva Health Forum, selected speaker presenting on "Participatory approaches and open data on venomous snakes: a neglected opportunity in the global snakebite crisis?", Geneva, Switzerland (April 10)

PUBLICATIONS IN PEER-REVIEWED SCIENTIFIC JOURNALS

2021	Martani A, Geneviève LD, Elger BS, Wangmo T. "It's not something you can take in your hands". Swiss experts' perspectives on health data ownership: an interview-based study. In press at BMJ Open.
2020	Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. BMC Med Ethics. 2020;21(1):17.
	Martani A, Geneviève LD, Poppe C, Casonato C, Wangmo T. Digital pills: a scoping review of the empirical literature and analysis of the ethical aspects. BMC Med Ethics. 2020;21(1):3.
2019	Geneviève LD, Martani A, Mallet MC, Wangmo T, Elger BS. Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: A systematic review. PLoS One. 2019;14(12):e0226015.
	Martani A, Geneviève LD, Pauli-Magnus C, McLennan S, Elger BS. Regulating the Secondary Use of Data for Research: Arguments Against Genetic Exceptionalism. Front Genet. 2019;10:1254.
	Geneviève LD, Martani A, Wangmo T, Paolotti D, Koppeschaar C, Kjelso C, Guerrisi C, Hirsch M, Woolley-Meza O, Lukowicz P, Flahault A, Elger BS. Participatory Disease Surveillance Systems: Ethical Framework. J Med Internet Res. 2019;21(5):e12273.
2018	Geneviève LD, Wangmo T, Dietrich D, Woolley-Meza O, Flahault A, Elger BS (2018) Research Ethics in the European Influenzanet Consortium: Scoping Review. JMIR Public Health Surveill 2018;4(4):e67.

Geneviève LD, Ray N, Chappuis F, Alcoba G, Mondardini MR, Bolon I, Ruiz-de-Castaneda R. (2018) Participatory approaches and open data on venomous snakes: A neglected opportunity in the global snakebite crisis? PLoS Negl Trop Dis 12(3): e0006162.

MANUSCRIPTS UNDER REVIEW/ SUBMITTED IN PEER-REVIEWED SCIENTIFIC JOURNALS

2021

Geneviève LD, Martani A, Perneger T, Wangmo T, Elger BS. Systemic fairness for sharing health data: perspectives from Swiss stakeholders. Under review at Frontiers in Public Health.

Geneviève LD, Martani A, Wangmo T, Elger BS. Structural Racism and Precision Public Health in the COVID-19 pandemic. Under review at the International Journal of Environmental Research and Public Health.

Geneviève LD, Martani A, Elger BS, Wangmo T. Individual notions of fair data sharing from the perspectives of Swiss stakeholders. Submitted to BMC Health Services Research.

BOOK CHAPTER

2019

Haesen S, Geneviève L, Elger B. Personnes agées en Prison. In: Santé en Prison. RMS Editions. Chêne-Bourg, pp. 317-330.