Conceptualizing Fairness in the Secondary use of Health Data for Research: a Scoping Review

Patricia Cervera de la Cruz[a*] Mahsa Shabani[a,b]

[a]*Metamedica, Faculty of Law and Criminology, University of Ghent, Belgium;* [b]*Law Centre for Health and Life, Faculty of Law, University of Amsterdam*

Correspondence details: Name: Patricia Cervera de la Cruz; Address: Metamedica, Faculty of Law and Criminology, Ghent University, Ghent 9000-B, Belgium; Email address: patricia.cerveradelacruz@ugent.be, ORCID: https://orcid.org/0000-0002-4869-4457

Journal name: Journal of Accountability in Research

Conceptualizing Fairness in the Secondary use of Health Data for Research: a Scoping Review

Abstract

With the introduction of the European Health Data Space (EHDS), the secondary use of health data for research purposes is attracting more attention. Secondary health data processing promises to address novel research questions, inform the design of future research and improve healthcare delivery generally. To comply with the existing data protection regulations, the secondary data use must be fair, among other things. However, there is no clear understanding of what fairness means in the context of secondary use of health data for scientific research purposes. In response, we conducted a scoping review of argument-based literature to explore how fairness in the secondary use of health data has been conceptualized. A total of 35 publications were included in the final synthesis after abstract and full-text screening. Using an inductive approach and a thematic analysis, our review has revealed that balancing individual and public interests, reducing power asymmetries, setting conditions for commercial involvement, and implementing benefit sharing are essential to guarantee fair secondary use research. The findings of this review can inform current and future research practices and policy development to adequately address concerns about fairness in the secondary use of health data.

Keywords: fairness; data sharing; secondary use; health data; health research

Introduction

The vast amount of health data that is generated nowadays has unlocked the door to a wide range of secondary uses for research, healthcare delivery, and policy making. An example of secondary uses for research purposes is the use of electronic health records that are originally collected and used to provide primary care to patients in the healthcare context for scientific research purposes,

such as facilitating antimicrobial surveillance in several countries (EHDS2 Pilot 2023) or revealing environmental factors that impact the expression of genetic variants (McCarty et al. 2011). While evidence has shown benefits of secondary use of health data in better addressing health issues globally, many technical barriers such as interoperability, poor data management, and the lack of standardized anonymization techniques are said to impede the realization of this potential (TEHDAS (Joint Action Towards the European Health Data Space) 2022a).

Furthermore, the sensitivity of health data has raised many legal and ethical concerns such as issues related to discrimination and bias (Ledford 2019; Zou and Schiebinger 2018), the implications for relatives or ethnic groups (Ross, Iguchi, and Panicker 2018), the potential for data misuse by third parties (O'Doherty et al. 2016) and privacy-associated risks such as re-identification (Rocher, Hendrickx, and de Montjoye 2019; Sweeney 2015) which have required adopting a robust ethics and regulatory framework. Among others, the secondary use of health data for research purposes is regulated in the European Union by the General Data Protection Regulation (GDPR) (*Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance)* 2016). While the reuse of data for scientific research is not in principle incompatible with the GDPR (Article 5 (1) (e)) the implementation of the Regulation in some aspects such as legal bases for secondary uses for research purposes has been fragmented across the Member States. This has led to the creation of what some have called "(…) a divergent and fragmented [regulatory] landscape." (Slokenberga 2020), with researchers struggling to gain access and share health data across borders (Boyd et al. 2021; Hansen et al. 2021). Another problem is that the 'secondary use of data' is not explicitly defined in the GDPR and has caused confusion regarding whether it constitutes 'further processing' under the GDPR or not (Becker et al. 2022). All of these issues have held back

developments in secondary use research, with some even claiming that the lack of a common framework for the secondary use of health data has posed significant barriers to Covid-19 research (Terzis 2022; McLennan, Celi, and Buyx 2020).

Looking to address some of these issues, the European Commission launched its proposal for a European Health Data Space (EHDS) (*Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space* 2022) in 2022. Among its intentions is to facilitate the secondary use of health data for a number of purposes, including for scientific research (Article 34 EHDS). The ambitions of the proposal and its expected impact for enabling the reuse of health data has provoked a wave of excitement and anticipation in Europe. However, alongside this excitement, and in light of the proposal, questions have been raised about whether the secondary use of health data is compatible with the overarching principle of fairness, which has been the cornerstone of the right to data protection as enshrined in Article 8 of the Charter of the Fundamental Rights of the European Union (European Union 2012) and identified as one of the guiding principles of data protection under the GDPR (Article 5 (1) (a)) Therefore, understanding what fairness means in the secondary use of health data is crucial to ensure compliance with the existing data protection legislation. Moreover, demonstrating that health data is processed fairly is a necessary step to ensuring trust, which has been identified to be a "key foundation" (Schinas 2023) of the EHDS proposal. Reaching some conclusions on which practices are considered to be fair may serve to be useful in guiding future secondary use research with health data and also contribute to address the existing ambiguity about what fairness means in this context (Slavkovik 2023).

In the context of the secondary use of health data for research purposes there is currently no clear understanding of what fairness is. While empirical studies have attempted to elucidate this

concept by exploring the views of individuals, (Kalkman et al. 2022) there remains unclarity regarding the fairness justifications and theoretical framework around it. Generally, the concept of fairness appears to be context-dependent and is often perceived differently, as what may be fair to some may not be fair to others (Terzis 2020). Therefore, while it may seem fair to some that individuals share their health data for research that will contribute to scientific knowledge generally, to others sharing their health data for research may only be fair where it is directly beneficial to their wellbeing. Having a better understanding of what is fair in secondary use research is therefore crucial to ensure the responsible and acceptable use of health data. In the absence of this conceptualization, researchers could fail to gain the trust needed from individuals to share their health data for secondary use research.

It is also necessary at this stage to make an important distinction between fair data sharing, conceptually understood, and FAIR data, which is data that follows a number of technical and organizational principles. These principles offer practical guidance for optimal data sharing and re-use, maintaining that data should be easily Findable, Accessible, Interoperable and Reusable, but does not include the concept of fairness as such. Against this backdrop and in the context of research more generally, the issue of fairness to researchers has been explored. These studies have revealed researchers expectations about fairness in data sharing, such as issues related to data access costs, recognition for contributions or compensation (Geneviève, Martani, Elger, et al. 2021; Geneviève, Martani, Perneger, et al. 2021; Devriendt, Shabani, and Borry 2021; Colledge and Elger 2015). These issues are important to protect researchers against unfair practices which limit equal opportunities for researchers. Nevertheless, they are conceptually different to what individuals may reasonably find to be fair or unfair in secondary use research.

Therefore, here we seek to understand what is understood to be *fair* in the secondary use of health data for research purposes.. To achieve this, we performed an up-to-date scoping review of the argument-based literature to explore the existing arguments relating to how fairness is conceptualized in the secondary use of health data. The results of this review form the basis for further reflections and discussions on the understanding of fairness in the secondary use of health data.

Methods

The scoping review framework was selected as a suitable approach for our goals in mapping the existing research studying conceptions of fairness in the context of secondary use of health data, rather than to analyze more specific research questions. Indeed, the guidance in the literature suggests the scoping review methodology for the purpose of clarifying "working definitions and conceptual boundaries in a topic or a field" (Peters et al. 2015). Moreover, to conceptualize fairness, it was decided that a focus on argument-based literature would be most suitable given that it enables a comprehensive overview of the arguments underpinning the conceptualization of fairness in this context.

Data sources

The databases of PubMed (MEDLINE), Web of Science and Embase were queried for articles. In addition, the first 15 pages of Google Scholar (150 records) were explored to complement the search. This cut-off was manually chosen due to the number of records identified by Google Scholar (>15,000). The following search string was used for the queries: (("secondary use" OR "reuse" OR "further processing" OR "data processing" OR "data sharing") AND "health data" AND (fair* OR "public interest" OR ethic* OR equit* OR benefi*)) (Box 1). The search was

carried out between January and March 2023. No limits on publication dates were used and a relevance-based search was conducted. The reference lists of selected articles were consulted to retrieve and include additional relevant articles.

Study selection

Articles were included in the scoping review if they complied with the following criteria: (1) include articles which discuss fairness (or related concepts) in the context of secondary use of health data for research purposes; (2) only discuss fairness from individuals' perspective; (3) are argument-based and (4) were published in peer-reviewed journals in English. Articles were excluded if they addressed fairness from a researcher, health practitioner or professional perspective and when they used empirical methodology. After selection based on title and abstract screenings, P.C. independently evaluated the full texts to determine their suitability for inclusion. In cases of uncertainty, articles were revisited by P.C. and M.S. and differences in interpretation discussed to reach consensus**.** The PRISMA flowchart details the study selection process (Figure 1).

Data extraction and synthesis

Quotes were extracted from the text and coded, and the extracted codes were discussed and verified by P.C. and M.S. Preliminary coding was conducted by P.C and the final codes were determined after discussions between both authors. Excel sheets were used to manage the coding. The results were analyzed using an inductive approach and thematic analysis (Braun and Clarke 2023), which is appropriate for the procedure of a scoping review. (Thomas et al. 2017) The structure of the categories was developed independently by both authors and discussed, and subsequently, a single thematic structure was developed.

Box 1. Key search terms
(("secondary use" OR “reuse” OR "further processing" OR "data processing" OR "data sharing") AND "health data" AND (fair* OR "public interest" OR ethic* OR equit* OR benefi*)) Asterisks (“*”) are used as a wildcard to allow any given search terms to be truncated or remain the same.

Results

Our scoping review identified 35 argument-based articles that met the inclusion criteria of the study (Table 1). The selected articles were analyzed and relevant quotes retrieved. Subsequently, the analyses of the extracted quotes led to the development of four main factors that impact the understanding of fairness in secondary use research: balance of interests, power asymmetries, commercial involvement and benefit sharing. The themes uncover the main building blocks for the conception of fairness in the secondary use of health data for research purposes.

Balance of interests

The literature reveals an existing tension between appeals to individual interests and public interest for the secondary use of health data for research. Thus, while to some authors, appealing to a public interest is a fair justification for the reuse of health data (Ballantyne and Schaefer 2020; Johnson, Kollnig, and Dewitte 2022), other authors believe that the loss of individual control or autonomy by participants is unfair (McCoy, Joffe, and Emanuel 2020; M. J. Taylor and Whitton 2020; Kaplan 2016).

Public interest

Those authors who appeal to *public interest* justifications argue it is fair to override individual rights and interests when the secondary use of health data is in the public interest (Ballantyne 2019; Townend 2018), especially in countries where individuals are said to benefit from a public healthcare

system. Indeed, there are even some authors that appeal to a moral duty to participate in research, especially given the minimal risks and high benefits that health data research produces (Ballantyne and Schaefer 2020; Jones et al. 2017). Therefore, some see it as fair to impose a duty to participate in the secondary use of health data for research purposes. Moreover, given the potential of secondary uses of health data to advance scientific knowledge, and therefore ultimately, improve the health of society (which is understood to be in the public interest), some authors maintain that it is fair to reuse this health data (Hepgul et al. 2019; Shabani 2022).

However, others take issue with the fact that while, in theory, public interest may be a good measure of the fairness of the secondary use of health data, in practice, it is not a well-defined concept (Grewal and Newson 2021; Tangcharoensathien, Boonperm, and Jongudomsuk 2010). Given that public interest has broad conceptions, this may pose a problem for conceptualizing fairness, given that the fairness in secondary use of health data may rest on an unspecified public interest. In the context of Covid-19, Ferretti and Vayena describe governments interpretation of the public's goodwill to share health data for the common good "as a free pass to reuse these data as they wished"(Ferretti and Vayena 2022). Looking to the example of health data that is collected by wearable devices such as step-counters, Meszaros and Ho question how far the concept of public interest can be stretched to use these types of health data.[22] Moreover, Taylor and Whitton argue that "(…) it does not follow that unfettered access to personal data for health research purposes is in the public interest. It is not the only common interest at stake." (M. J. Taylor and Whitton 2020) Therefore, there seems to be an agreement even by those who defend the public interest as a justification for the secondary use of health data that this justification cannot be boundless. Other authors extend this concern also to the use of a public interest justification by commercial actors to gain access to health data (Cheung 2020). Nevertheless, some authors defend

the use of a broad definition for public interests, claiming that it is in fact this broad understanding that makes it particularly agile and apt for considering a wide range of relevant interests, particularly "given that progress in science is rarely linear and technology within health is often transferable between different patient populations." (Ballantyne and Schaefer 2021)

In defending the primacy of public interests over individual interests, some authors suggest that refusing to share health data for research purposes is unfair and unsolidarious, particularly in countries where a public healthcare system is in place (Ballantyne and Schaefer 2020; Hepgul et al. 2019; Meszaros and Ho 2019). The argument here is that individuals that benefit from the healthcare system should contribute to it by sharing their health data for research. Despite the recognition that access to public health services are generally inequitable among different population groups (Ballantyne and Schaefer 2020), some maintain that everyone enjoys at least some benefits (M. J. Taylor and Whitton 2020; Ballantyne and Schaefer 2018), such as sewerage or sanitation, from these services and therefore "(…) it is an overstatement to suggest that many people fail to benefit at all from health knowledge, or even benefit very little." (Ballantyne and Schaefer 2018) Moreover, Taylor and Whitton also argue that the fact that we are all vulnerable and our future is uncertain means that we cannot confidently assume that secondary use research will never be individually relevant to us (M. J. Taylor and Whitton 2020). Understood this way, the authors argue that the secondary use of health data for research purposes is in the public interest since the uncertainty of our own health and our shared vulnerability means that we may all reap the benefits at some point in time.

Individual interests

Concerned with the impact of the secondary use of health data on *individual interests*, some authors claim that the re-use of health data for research purposes is unfair if patients lose the sense of control or autonomy over their health data. Indeed, having the ability to exercise some control over

individuals' own data is seen as an important element in determining how fair a health data 'exchange' is. Control over health data can take many forms, from making decisions about how the data will be reused, to having the ability to provide consent (and withdraw it) or to limiting access to certain actors. So, for instance, having the ability to opt-out at any point from research, even when this is in the public interest, is seen as key to ensuring that this research is fair individuals (Hepgul et al. 2019). Therefore, the inclusion of an opt-out may strike a good balance between the need for individuals to retain control over their health data and the public interest in having access to large amounts of patient health data. However, Meszaros and Ho underline that "(I)f citizens are empowered to exercise control over their data, it is crucial to clarify the limits of their decisions (…)" (Meszaros and Ho 2019).

To some, this loss of control is seen as insignificant, since in comparison to other types of research, health data research does not involve an invasion of a person's bodily integrity and therefore "(…) does not directly affect any individual's personal experiences or life course (…)" (Ballantyne and Schaefer 2020). In fact, any loss of control or of individual rights over health data in the case of secondary use of health data for research purposes is sometimes even regarded as a fair "trade-off" (M. J. Taylor and Whitton 2020; Cheung 2020). However, others believe this trade-off is a "(…) fallacy (…) used by institutions to legitimate widespread collection of personal data." and is unfair because most of the times individuals are not fully made aware of the consequences of giving up their data (Cheung 2020). Moreover, it is argued that conceiving the secondary use of health data for research as a trade-off is unfair as it gives the impression that that individuals are instrumentalized or used as a means to an end (M. J. Taylor and Whitton 2020). All in all, the widespread agreement in the literature is that researchers must make clear what the reasons or rationales for a particular trade-off are and avoid making the assumption that it will be acceptable

to all the individuals (Shabani 2022; Tangcharoensathien, Boonperm, and Jongudomsuk 2010; Meszaros and Ho 2019; Cassel and Bindman 2019).

Overall, the results show that there is a feeling of irreconcilability between those who view a fair secondary use of their health data as dependent on the level of individual control they have and those who view it as dependent on its potential to benefit society. Reflecting on this issue, Townend claims that "(…) there is an underlying difficulty in appealing to simple solidarity to many who embrace more individualistic autonomy" (Townend 2018).

Power asymmetries

The existing power imbalances or power asymmetries between individuals (data subjects) and those processing the data (data controllers) are also identified as a barrier for ensuring the fair secondary use of health data (Ploug and Holm 2017; Prainsack et al. 2022; Sharon 2018). While in traditional or standard research it is usually clear to tell who is using the data and for what purposes, in the secondary use of health data for research, individuals may not even be aware of their data being used (Cassel and Bindman 2019; Kaplan 2016), making it difficult to identify who is processing their health data in the first place.

While, as discussed above, some authors suggest that giving individuals certain levels of individual control over their health data is sufficient to remedy these concerns, Prainsack et al. maintain that "(…)strengthening the control that individuals have over their data are not enough" (Prainsack et al. 2022) since the power asymmetries that exist between data subjects and data processors are too significant. In the same line, Shabani argues that individuals have limited "(…) negotiating power when it comes to how their data are to be shared, under what conditions, and how the final outcomes of the data use are to be distributed." (Shabani 2022) On that account, many

contend that the involvement of large commercial actors, such as Big Tech, in secondary use research is unfair given their huge influential power (Ferretti and Vayena 2022; Cheung 2020; Marelli, Testa, and Van Hoyweghen 2021).

The existence of these power imbalances has been argued by some to have potential repercussions on individuals' trust in research. Ballantyne and Schaefer contend that a decrease in trust may lead some individuals "(…) to withdraw from health services which are contributing to data, or be less willing to participate in voluntary studies (…)" (Ballantyne and Schaefer 2018).

Commercial involvement

The involvement of commercial interests in the secondary use of health data for research is repeatedly identified as a main concern for fairness towards individuals. While some authors consider it to be unfair because commercial involvement shifts the focus from improving healthcare to achieving financial profits, others argue that limiting commercial or private contributions to scientific research may pose a significant limitation for the secondary use of health data for research.

Financial interests and profit

It is generally regarded to be unfair when private or commercial entities receive large profits from health data research. Some authors argue that when profits are involved, there is a risk that "(I)ndividuals might rightfully feel taken advantage of if they suspect that their health information is being monetized, without their consent (…)" (Cassel and Bindman 2019). In fact, when individuals are particularly vulnerable, the practice may even be deemed exploitative (McCoy, Joffe, and Emanuel 2020). Given that the most useful health data for research will be that of sick patients who have the most complete and extensive electronic health records (Johnson, Kollnig, and Dewitte 2022),

Ferretti and Vayena warn that the risk that the private sector will be "(…) exploiting sick people for corporate profit must be considered carefully" (Ferretti and Vayena 2022).

However, other authors maintain that private entities that conduct research on patient data " (…) are entitled to earn fair profits from their investments" (McCoy, Joffe, and Emanuel 2020). Therefore, some see it as unfair to place a large financial burden on private companies, especially since this may disincentivize them to develop products and treatments that are potentially beneficial to the general public (McCoy, Joffe, and Emanuel 2020). Moreover, the involvement of private actors, that is non-public bodies, is also considered to be key to innovating the healthcare sector and therefore in the public interest (Ballantyne and Schaefer 2018; Sharon 2018). Indeed, some authors recognize that the involvement of these private actors may be necessary to reach the expected benefits of the secondary use of health data and that we should therefore accept some risks associated to their involvement, like potential misuses of data, for the sake of the potential benefits they may bring (Sharon 2018; Seastedt et al. 2022).

Furthermore, Ballantyne and Schaefer argue that the fact that private actors may receive large financial profits is not a concern since the so-called "economic trickle-down effect to the public health sphere" (Müller 2022) will apply, meaning that expensive medical products or treatments available only to the wealthy, will trickle down to other social classes. Yet, the lack of evidence for this, especially in the healthcare sector, leads other authors to believe that the accessibility of the products resulting from the secondary use of health data is a central issue that needs addressing (Tangcharoensathien, Boonperm, and Jongudomsuk 2010; Ramsay 2022).

Nevertheless, some authors argue that obliging commercial companies to return financial benefits to communities may do more harm than good and that it is unfair to impose this type of financial obligation on them (McCoy, Joffe, and Emanuel 2020). These financial pressures may push

companies to withdraw from health research due to its high costs, hindering future innovation in healthcare. Sharon argues that it is in fact this entrepreneurial drive that can be a real asset in secondary use research, arguing that "these companies have the technological expertise required to achieve scientific breakthroughs" (Sharon 2018). Furthermore, some authors argue it may be unfair to demand that all secondary research guarantees benefit-sharing since "(…) specific research goals are not (and cannot) always be established (…)" (Ganguli-Mitra 2012) a priori. Since secondary use research is so unpredictable, anticipating the benefits that may result and how to distribute these equitably may prove to be very difficult, if not impossible. Yet, others argue that precisely given this uncertainty about benefits to patients, "data-sharing agreements should take proactive measures to ensure that the financial benefits of data sharing are fairly distributed" (McCoy, Joffe, and Emanuel 2020).

Intellectual property and access limitations

The issue of intellectual property rights poses a particular challenge to fairness with regards to commercial involvement in secondary use health data research. Placing intellectual property rights or other mechanisms such as paywalls is understood to be unfair because it "(…) can shift benefits into private spheres (…)" (Müller 2022). To illustrate the actual risk to accessibility that the involvement of private actors in health research may pose, Cheung makes reference to DeepMind's (subsequently Google) collaboration with London's Moorfield Eye hospital (Cheung 2020). In this partnership, the hospital transferred identifiable patient records to DeepMind for the development of an AI on eye scans to detect early symptoms of sight loss. Critically, the AI was made available to use by the hospital for the first five years but the future costs associated to the use of the product were not detailed. Therefore, any restrictions placed on the access of products resulting from the

secondary use of health data are considered to be unfair since it may further existing inequities in healthcare, making those most vulnerable worse-off (Golinelli et al. 2018).

Influence on research agenda

Some argue that it would be unfair to involve private actors in the secondary use of health data since there is a risk that they will monopolize research and set their own research agendas, which may not be aligned with the interests of individuals (Cheung 2020). Talking about Big Tech platforms, Marelli et al. argue that "they are increasingly poised to become obligatory passage points in the digitized biomedical landscape of the coming decades" (Marelli, Testa, and Van Hoyweghen 2021). Indeed, as put forward by Cheung, Apple's ResarchKit Tool, which constrains researchers to use Apple devices to conduct their research, is an example of the risk of monopolization of the research market by Big Tech companies (Cheung 2020). This not only limits competition in the research sector, but creates further health disparities since "only a specific socio-economic demographic—owners of iPhones—can participate in ResearchKit studies" (Sharon 2018). Thus, to these authors it seems unfair to allow those most powerful to dictate how research is conducted, in part because of their track-record of "(…) contributing to new patterns of global inequalities (…)" (Cheung 2020) and because of their potential to constrain the ways in which research can be conducted.

Furthermore, private actors are more likely to have "vested interests" (Bull, Roberts, and Parker 2015) and may therefore push for specific research agendas based on these. If secondary use research does "undermine patients' interests" (Johnson, Kollnig, and Dewitte 2022) by increasing a companies' profits for example, it may be unfair to allow private actors to conduct health data research. Nevertheless, these conflicts of interest may also present themselves in the public sector, such as in the case of the Danish circumcision registry where Ploug and Holm argue that "(T)here

may be a perceived conflict of interest between the perceived opposition to circumcision and the ability to independently and objectively evaluate and interpret evidence about its safety (…)" (Ploug and Holm 2017).

Despite acknowledging the reality that commercial actors often "mobilize the language of public good, altruism and solidarity" to exploit patients, Sharon argues that "'doing good' is becoming an inalienable—not an additional—dimension of corporate activity" (Sharon 2018). Therefore, secondary use research may in fact be more than purely a financial venture for some commercial actors and may also involve good intentions. From this perspective, it may seem unfair to exclude private actors on a purely reputational basis. Moreover, Marelli et al. also contend that it is increasingly becoming "(…) difficult to neatly discriminate between undertakings that primarily pursue collective knowledge and wellbeing as opposed to private interests (…)." (Marelli, Testa, and Van Hoyweghen 2021)

Benefit sharing

Many argue that benefits in secondary use research should not come as an after-thought. Ramsay argues that we have a "(…) duty to the communities and research participants, as custodians of their data, to remain mindful of the need to carefully consider the nature of potential benefit." (Ramsay 2022) It is generally agreed that a pre-condition to achieving health equity is that there is "a fair distribution of health outcomes in societies." (Ballantyne 2019) Therefore, benefit sharing is seen as a step further to the balancing of risks and benefits, focusing on the equitable distribution of these (Murtagh et al. 2021).

Taxation and distribution of financial profits

Taxation has been suggested to be a way to implement benefit-sharing, especially when private actors are involved. In fact, Prainsack et al. point to the European Commission's proposal to introduce corporate taxation in the digital economy and argue that it could "(…) help to reduce inequities between countries where digital businesses operate and where they pay taxes." (Prainsack et al. 2022) This tax should only be imposed on health data generated in the public domain and should therefore be redistributed for "public needs in the health sector" (Sharon 2018). Moreover, Prainsack et al. argue that "(W)here commercial profits accrue from data use, some of these profits need to come back to the people and communities that enabled them in the first place." (Prainsack et al. 2022) Similarly, Johnson et al. claim that "(I)f commercial actors may benefit from access to health data, there ought to be similar benefits for patients, research participants and their communities, in exchange (…)" (Johnson, Kollnig, and Dewitte 2022). Therefore, the return of profits from research undertaken by commercial actors to the communities is seen as important in achieving fairness in the secondary use of health data. Indeed, Ganguli-Mitra argues that in contrast to the traditional medical research model in which the presumption of therapeutic benefits was sufficient justification, in the case of secondary use research, the advent of financial profits requires researchers to rethink how and if they should be distributed with the data subjects (Ganguli-Mitra 2012). Marelli et al. also make a similar analogy, claiming that benefit sharing must go beyond the "post-study obligations of clinical studies" and instead focus on avoiding the "depletion of public assets and technological know-hows, especially when publicly-collected data is transferred to powerful entities such as Big Tech platforms." (Marelli, Testa, and Van Hoyweghen 2021) Therefore, benefit sharing in the secondary use of health data for research involves a radical shift into how benefits are distributed with patients.

Fontana et al. argue that where public data can deliver direct financial value, it is the responsibility of the public sector to recognize this value and return this to the taxpayer, claiming that a failure to do so will likely lead a negative public reaction and diminished trust in the competence of public healthcare administration (Fontana et al. 2020). Thus, some argue that if the profits from the sale of health data are reinvested into the public healthcare system, whose "(…) economic sustainability (…) represents a concrete and current challenge for governments (…)" (Golinelli et al. 2018), it would be fair to permit it, given that it would serve the public interest (Johnson, Kollnig, and Dewitte 2022). Nevertheless, even where this is the case, Johnson et al. maintain that the "(…)the primary beneficiaries ought to be the patient communities" (Johnson, Kollnig, and Dewitte 2022). Therefore, it is seen as fair for the public health sector to seek to gain financial value in individuals' health data, so long as this is reinvested to improve existing health services. However, Fontana et al. also recognize that some public healthcare systems, such as the National Health Service (NHS) in the UK, do "(…)not currently have the resources and skills to capture the value of its data" (Fontana et al. 2020) and may therefore need to seek out help from the private sector to achieve these goals.

Data commons

Data commons are also put forward by some authors as a way to achieve fairer outcomes in secondary use research and equitable distribution of these among communities (Prainsack et al. 2022; Sharon 2018; Kaplan 2016). Sharon defines data commons as "(…) self-governing communities and cooperatives (…) whereby collective decision-making, reciprocity and shared benefit are foregrounded, and the social value of health data is prioritized over its economic value." (Sharon 2018) This type of "solidarity-based governance", as described by Prainsack et al., is considered to make sure that benefits and burdens are experienced equitably by all data subjects involved

(Prainsack et al. 2022). The engagement of communities in the decision-making related to their health data use is considered to render the secondary use of health data fairer and, as a result, more trustworthy (McCoy, Joffe, and Emanuel 2020; Murtagh et al. 2021). Nevertheless, some authors question whether it is reasonably "(…)fair to expect this level of engagement on the part of citizens (…)" (Sharon 2018).

Affordable access to benefits

The affordable access to the benefits from secondary use research is identified as another key component to ensure fairness.. Gross et al. argue that the availability of therapeutic devices and procedures developed by government-funded research is, in effect, already mandated by the Belmont Report which demands that "these [do] not provide advantages only to those who can afford them" (Gross et al. 2022). In the context of Low- and Middle-Income Countries (LMICs) this has been repeatedly recognized as a problem, revealing the unfair research practices individuals experience. Tangcharoensathien et al. recall how the "(…) sharing of avian flu virus specimens by developing countries through the World Health Organization resulted in the production of avian influenza vaccines at a price of US$ 10–20 per dose.", (Tangcharoensathien, Boonperm, and Jongudomsuk 2010) a price tag that made it entirely unaffordable in low-income countries. Speaking from the African perspective, Ramsay also argues that "(W)e cannot accept a status quo that limits access to novel applied technologies (…) only to those who can afford it" (Ramsay 2022). Knowing "the historical pattern" (Gross et al. 2022) in which health research has typically failed to provide affordable access to therapies and treatments, it seems unfair to ask patients to trust that the secondary use of their health data will result in tangible benefits to them. It is argued therefore that the secondary use of health data must be " (…) subjected to the burden of proof." (Correia, Rego, and Nunes 2021) if they claim to be beneficial to data subjects. Similarly, benefits need to be tangible

and visible to communities because "(I)t is important to have evidence of the benefits that populations receive directly as a result of sharing, beyond publications by secondary users." (Tangcharoensathien, Boonperm, and Jongudomsuk 2010)

A suggestion proposed by Marelli et al. to address these issues of accessibility and "fairly balance the distinct values and interests at stake in health research." (Marelli, Testa, and Van Hoyweghen 2021) is to offer those products which are developed by processing publicly held data free-of-cost. Although, Sabatello et al. acknowledge that setting these types of conditions for the secondary use of health data may add complexity, they also argue that it "facilitates changes towards socially responsible perceptions of the public good" (Sabatello et al. 2022). Protecting outcomes of secondary use research from paywalls and intellectual property rights is therefore seen as crucial to securing a fair secondary use of health data.. In the special context of LMICs, authors demand greater accessibility to benefits for "those most affected by poverty, poor health, and low access to resources." (Ramsay 2022)

Capacity building

Some authors, especially those concerned with the fairness of secondary use of health data in LMICs, argue that ensuring capacity building is a key requirement to ensure that the secondary use of health data is sufficiently fair in those countries (Tangcharoensathien, Boonperm, and Jongudomsuk 2010; Ramsay 2022; Bull, Roberts, and Parker 2015; Ganguli-Mitra 2012). Tangcharoensathien et al. present an example of capacity building in Thailand, claiming that "the National Statistical Office grants approval to international secondary users to access nationally representative household data sets, on the condition that they develop partnerships with local scientists." (Tangcharoensathien, Boonperm, and Jongudomsuk 2010) Similarly, Bull et al. argue that the "(P)rinciples of fairness and reciprocity also suggest that resources be made available to enable researchers in low- and middle-

income settings to undertake secondary analysis of data sets that have the potential to address their research priorities." (Bull, Roberts, and Parker 2015)

Individual level benefits

Cassel and Bindman suggest that financial benefits could be distributed with individuals by way of a license "that results in royalty payments every time their data are included in a research study" or by offering an equity distribution for products that result (Cassel and Bindman 2019). However, some authors express concerns regarding the direct payment to data subjects for their health data, claiming it may be exploitative (Ganguli-Mitra 2012). This risk may be particularly pronounced in LMICs therefore demanding that benefit sharing frameworks pay particular attention to the context in which they are established (Ganguli-Mitra 2012). Others also claim that "(S)haring financial benefits of data sharing with communities rather than individuals may be a more effective approach to ensure a fair distribution of benefits."(McCoy, Joffe, and Emanuel 2020) since health data is only valuable when it is aggregated.

Furthermore, and especially where individuals are concerned with the individual benefits that they may expect from participation in secondary use research, the timeliness of therapeutic benefits may be crucial in determining whether this is fair to them or not. In the case of secondary use research with certain patient groups, for example, individuals may expect to receive improved care and treatment for their illness as a result of their participation. Indeed, where results are timely and actionable, some authors argue that it is fair to guarantee the return of results to individuals (Ganguli-Mitra 2012; Francis and Francis 2017). Not doing so, Gross et al. argue, would "appear to trump the fiduciary duty to benefit an individual with the products of their own care" (Gross et al. 2022).

Discussion

In this paper, we set out to explore the conceptualizations of fairness in the secondary use of health data for research purposes. The analysis of the articles included in our review revealed four main factors that impact the understanding of fairness in secondary use research: balance of interests, power asymmetries, commercial involvement and benefit sharing. Notably, our review highlights that there is a need to find an appropriate balance between individual interests and public interests in secondary use of health data; that the perceived asymmetry of information that exists between data users and data subjects in health data reuse should be addressed; that the involvement of commercial entities raises concerns regarding financial profits, their return and influences on research agenda setting; and finally, that it is important that accrued benefits, be they financial or non-financial, are appropriately shared with the individuals.

Our findings highlighted the importance of striking a balance between public and individuals' interest in secondary uses of health data. Although public and private interests are often put in contrast against each other in the literature, one may argue that this is misleading as there are many interests involved which may be shared by both the public and individuals (Sharon 2018; Prainsack et al. 2022; Prainsack 2018). Indeed, the Covid-19 pandemic was one example manifesting our shared interest in health protection in the context of a public health emergency. Moreover, not all agree that public interest in secondary uses of health data for research purposes would justify removal of all control on data by the individuals (Meszaros and Ho 2019; Cheung 2020; M. J. Taylor and Whitton 2020). Our review has shown that there is a lack of consensus regarding how to maintain this control in practice in the secondary use of health data, and questions remain on the extent to which the protection of individual interests can be secured through individual control mechanisms such as consent (Erikainen et al. 2021).

Looking to find a compromise, the suggestion to use an opt-out mechanism has been put forward in the literature and is even included in the European Parliaments' proposed amendments to the EHDS proposal of May 2022 ("Committee on the Environment, Public Health and Food Safety and Committee on Civil Liberties, Justice and Home Affairs. Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space.," n.d.). Opt-outs for the secondary use of health data for research are already in place in some countries such as England and Finland (NHS England 2023; Findata 2022), as part of national initiatives to facilitate the secondary use of health data. Another emerging concept, which is argued to enhance individual's control over their health data (Shabani and Yilmaz 2022), is data altruism consent. The concept, introduced in the Data Governance Act, (*Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA Relevance)* 2022) allows the voluntary sharing of data based on consent for uses in the public interest, such as for research purposes. In this way, individuals have the ability to choose to make their data available to trusted, not-for-profit organizations for the benefit of society. Data altruism consent allows individuals to give consent to the processing of personal data by data altruism organizations, which will be registered and recognized in the EU, for the objectives of general interest. Nevertheless, many remain skeptical of the value of the use of such consent models in this context, arguing that informed consent as it has been foreseen by data protection regulations requires more granularity and specificity (TEHDAS (Joint Action Towards the European Health Data Space) 2023a).

Additionally, our review revealed that when appeals to the public interest are made, it is important that what is in the public interest is well defined *a priori*. This lack of clarity regarding the public interest has been previously highlighted to be an issue in both the US in The Health Insurance Portability and Accountability Act of 1996 (HIPAA)(Kaplan 2016) and in the EU, in the

GDPR, where " (…) public interest is mentioned 70 times, yet on none of these occasions is the concept fully explained." (Slokenberga 2020) Aware of the ambiguity of the term, the advised approach in Australia's Data Availability and Transparency Act 2022 has been that a public interest assessment "should be made on a project-by-project basis, weighing the range of factors for and against sharing" (Commonwealth Parliament; Parliament House Canberra n.d.). Moreover, the Joint Action Towards a European Health Data Space (TEHDAS) reported that citizens' values should inform what constitutes the common good (TEHDAS (Joint Action Towards the European Health Data Space) 2023b). Public interest cannot be used as a catch-all term to allow all secondary use research and "(…) a social license for data-intensive health research cannot simply be presumed." (Kalkman et al. 2022) In light of the upcoming EHDS, where public interest is identified as one of the purposes for which electronic health data can be processed for secondary use (Article 34), it seems that future health data access bodies should strive to have a common and shared definition of what types of secondary use research are in the public interest. Although the discussion is ongoing, previous attempts to determine what is in the public interest, such as the framework developed in the UK by Involve, The Carnegie UK Trust's and Understanding Patient Data for evaluating public benefit (Involve, Carnegie UK Trust, and Understanding Patient Data 2018), can help to inform regulators and policy makers. The framework, established through a series of workshops with the public, identifies five criteria to determine the public benefit of a particular data sharing activity, including that data sharing benefits individuals' wellbeing, provides benefits for wider society, respects individuals and their privacy, represents an effective use of public resources and produces tangible benefits. Being explicit about the reasons why secondary use research is in the public interest is therefore key to ensuring it is fair.

Power asymmetries were also found to pose a significant obstacle for the fair secondary use of health data for research. Where individuals are not aware of their data being used for secondary purposes and where the data users are not clearly identified, this is argued to be unfair.. Individuals who are not adequately informed about the use of their data are more vulnerable and more susceptible to power imbalances (Malgieri and Niklas 2020). This asymmetry of information between individuals and data controllers has been previously highlighted (Chen, Dove, and Bhakuni 2022) and seems to be at odds with the right to information and the principle of transparency as they are mandated in the GDPR (Article 13 and Article 5 (1) (a)). Elsewhere in the literature, some even argue that the EHDS, as per the 2022 draft proposal, is creating new information asymmetries for individuals, due to the fact that there are no obligations on health data access bodies to inform natural persons concerning the use of their health data (Terzis 2022). Indeed, the EDPB-EDPS Joint Opinion on the EHDS Proposal note that the proposal restricts the right to information under the GDPR (European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS) 2022). To remedy this, many point to the importance of engaging individuals in research by providing clear and transparent communication about health data use (L. Taylor 2017; Vayena et al. 2018), ensuring individuals are adequately informed about their rights (Williams and Fahy 2019) or by giving individuals the ability to make decisions about their data use (Baker, Kaye, and Terry 2016). Involving individuals in the governance of secondary use research, by sitting in data access bodies or boards or participating in citizen-run health data collectives and cooperatives can also reduce power asymmetries and therefore ensure a fairer secondary use of health data (Blasimme, Vayena, and Hafen 2018). Furthermore, as discussed above, an opt-out may give individuals control over their health data and help to tackle these power asymmetries. However, to ensure the right to information and to protect individuals from power imbalances, it is important to notify individuals about their right to opt-out well in advance and share clear information about its limitations and

conditions for individuals to be able to make an adequately informed decision (Kuntsman, Miyake, and Martin 2019).

In close relation to the issue of power asymmetries, commercial involvement is another aspect that has an impact on the discussion about fairness. Notably, where large financial profits are accrued from secondary use research without any return to society, this is seen to be unfair.. Indeed, empirical research with individuals shows that commercial involvement is generally unwelcomed in secondary use research (Richter et al. 2021; The Wellcome Trust. 2017; Haddow et al. 2007). Nevertheless, these findings need to be reconciled with the reality that cooperation between public research institutions and commercial companies is commonplace and differentiating public from private interest is oftentimes difficult. (Marelli, Testa, and Van Hoyweghen 2021) Indeed, the European Commission's European Data Strategy recognizes the need to allow data access to commercial companies in order to achieve its goal of making "(…) the EU a leader in a data-driven society." (European Commission. 2019) Moreover, some qualitative studies reveal that individuals are more accepting of commercial involvement when they are involved in a deliberative process and receive transparent information about the uses of their health data by commercial actors (Tully et al. 2019; Street et al. 2021). However, it is important to understand what features about commercial involvement are unfair and look to address these elements. Our findings revealed that commercial involvement may be fair where certain conditions are met, such as benefit sharing (see below). Therefore, to guarantee research is fair, policy makers should ensure that large financial profits resulting from research are returned to individuals and society. This not only ensures that research is fair but it also helps to build and maintain trust in research (Williams and Fahy 2019; Waind 2020). Moreover, through the use of appropriate governance mechanisms, it is possible to limit the influence of large commercial actors on research agendas. Accountability measures, such as

imposing sanctions when data is misused, are also identified to be necessary to keep commercial actors in check (McMahon, Buyx, and Prainsack 2020). Therefore, in order to guarantee fairness in the secondary use of health data for research, it is important to develop concrete terms and conditions of use for commercial actors.

In particular, when commercial actors are involved in secondary use research, distributing the benefits of research equitably with individuals, or benefit sharing, is a crucial element in ensuring it is fair.. In the last twenty years, benefit sharing has become an essential element for research ethics (Dauda and Dierickx 2013) and has been recognized to enhance fairness (Munung and de Vries 2020). Indeed, as Ravinetto et al. argue, benefit sharing is stipulated in the Declaration of Helsinki claiming that "(…)if medical research is done within a vulnerable group, this group should stand to benefit from the knowledge, practices, or interventions resulting from the research." (Ravinetto et al. 2014) However, since some argue that the physical risks to individuals are rather insignificant in secondary use research compared to other types of research (Porsdam Mann, Savulescu, and Sahakian 2016), the importance of benefit sharing in this context has been disregarded. Nevertheless, our review shows that benefit sharing cannot be neglected in secondary use research. Not only does benefit sharing meet individuals' self-reported expectations about the use of their health data (Aitken et al. 2018; TEHDAS (Joint Action Towards the European Health Data Space) 2022b), but it ensures that the research is fair to them. The application of benefit sharing measures such as taxation (where profits are made from research), securing affordable access to the results from research, redistributing economic benefits with individuals and communities or ensuring capacity building are crucial in guaranteeing a fair secondary use of individuals' health data. Elsewhere, others have also noted the importance of making the results of research available to the general public (Ravinetto and Singh 2022; Winkler et al. 2023), something which is already required in the

EHDS proposal (Article 46). Moreover, qualitative studies with individuals and patients suggest that health data sharing is valued to be fair where it can to lead to faster development of therapies and improved treatment options for them (O'Brien et al. 2019; Lounsbury et al. 2021).

Limitations

A total of 35 peer-reviewed articles in English qualified for inclusion and were further assessed. It might therefore be possible that studies in other languages and relevant grey literature have been overlooked. Similarly, it is important that the included articles mostly consider secondary uses of health data from a Western perspective and therefore the results may not reflect broader, global views on the topic, resulting in a 'Westernized' conception of fairness. Moreover, the exclusion of empirical-based literature from the review may have overlooked other factors that impact fairness not discussed in the argument-based literature.

Conclusion

The secondary use of health data for scientific research is set to escalate within the European Union with the proposed EHDS. Given this, it is important that policy-makers adequately address emerging concerns, including concerns about fairness.. Based on the literature reviewed, balance of interests, power asymmetries, commercial involvement and benefit sharing are factors that impact fairness.. The first draft of the EHDS proposal seems to fail to address some of these issues, especially since citizens are afforded very limited decision-making power to decide how their health data is used. As we have demonstrated in this study, fairness demands a greater involvement of individuals in the secondary use of their health data. Policy makers should look to reconcile this call for greater individual control with the public interests involved in the EHDS. Moreover, a fair secondary use of health data needs to demonstrate a return of benefits for individuals so that no

one is left behind. Future research should look for ways to practically implement and operationalize fairness in the secondary use of health data for research purposes.

Declaration of interest statement

The authors report there are no competing interests to declare.

Funding

This work was supported by the Special Research Fund (BOF) at Ghent University and by PROPHET (A Personalized Prevention Roadmap for the Future Healthcare) which is funded by the European Commission under the Horizon Europe Research and Innovation Programme (grant agreement 101057721).

References

Aitken, Mhairi, Carol Porteous, Emily Creamer, and Sarah Cunningham-Burley. 2018. “Who Benefits and How? Public Expectations of Public Benefits from Data-Intensive Health Research.” *Big Data & Society* 5 (2): 2053951718816724. https://doi.org/10.1177/2053951718816724.

Baker, Dixie B., Jane Kaye, and Sharon F. Terry. 2016. “Governance Through Privacy, Fairness, and Respect for Individuals.” *eGEMs* 4 (2): 1207. https://doi.org/10.13063/2327-9214.1207.

Ballantyne, Angela. 2019. “Adjusting the Focus: A Public Health Ethics Approach to Data Research.” *Bioethics* 33 (3): 357–66. https://doi.org/10.1111/bioe.12551.

Ballantyne, Angela, and G. Owen Schaefer. 2018. “Consent and the Ethical Duty to Participate in Health Data Research.” *Journal of Medical Ethics* 44 (6): 392–96. https://doi.org/10.1136/medethics-2017-104550.

———. 2020. “Public Interest in Health Data Research: Laying out the Conceptual Groundwork.” *Journal of Medical Ethics* 46 (9): 610–16. https://doi.org/10.1136/medethics-2020-106152.

———. 2021. “In Defence of a Broad Approach to Public Interest in Health Data Research.” *Journal of Medical Ethics* 47 (8): 583–84. https://doi.org/10.1136/medethics-2020-106880.

Becker, Regina, Davit Chokoshvili, Giovanni Comandé, Edward Dove, Alison Hall, Colin Mitchell, Fruzsina Molnar-Gabor, Pilar Nicolás, Sini Tervo, and Adrian Thorogood. 2022. “Secondary Use of Personal

Health Data: When Is It 'Further Processing' under the GDPR, and What Are the Implications for Data Controllers?" SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4070716.

Blasimme, Alessandro, Effy Vayena, and Ernst Hafen. 2018. "Democratizing Health Research Through Data Cooperatives." *Philosophy & Technology* 31 (3): 473–79. https://doi.org/10.1007/s13347-018-0320-8.

Boyd, Mark, Dr Milly Zimeta, Dr Jeni Tennison, and Mahad Alassow. 2021. "Secondary Use of Health Data in Europe," 38.

Braun, Virginia, and Victoria Clarke. 2023. "Toward Good Practice in Thematic Analysis: Avoiding Common Problems and Be(Com)Ing a Knowing Researcher." *International Journal of Transgender Health* 24 (1): 1–6. https://doi.org/10.1080/26895269.2022.2129597.

Bull, Susan, Nia Roberts, and Michael Parker. 2015. "Views of Ethical Best Practices in Sharing Individual-Level Data From Medical and Public Health Research: A Systematic Scoping Review." *Journal of Empirical Research on Human Research Ethics* 10 (3): 225–38. https://doi.org/10.1177/1556264615594767.

Cassel, Christine, and Andrew Bindman. 2019. "Risk, Benefit, and Fairness in a Big Data World." *JAMA* 322 (2): 105–6. https://doi.org/10.1001/jama.2019.9523.

Chen, Jiahong, Edward S. Dove, and Himani Bhakuni. 2022. "Explicit Consent and Alternative Data Protection Processing Grounds for Health Research." *Research Handbook on EU Data Protection Law*, April, 474–502.

Cheung, Sarah. 2020. "Disambiguating the Benefits and Risks from Public Health Data in the Digital Economy." *Big Data & Society* 7 (1): 2053951720933924. https://doi.org/10.1177/2053951720933924.

Colledge, Flora, and Bernice Elger. 2015. “Getting a Fair Share: Attitudes and Perceptions of Biobank Stakeholders Concerning the Fairness of Sample Sharing.” *Bioethics* 29 (6): 424–30. https://doi.org/10.1111/bioe.12134.

“Committee on the Environment, Public Health and Food Safety and Committee on Civil Liberties, Justice and Home Affairs. Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space.” n.d. https://www.europarl.europa.eu/doceo/document/CJ43-PR-742387_EN.pdf.

Commonwealth Parliament; Parliament House Canberra. n.d. “Data Availability and Transparency Bill 2020 [and] Data Availability and Transparency (Consequential Amendments) Bill 2020.” Text. Australia. Accessed April 21, 2023. https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/bd/bd2122a/22bd018.

Correia, Mónica, Guilhermina Rego, and Rui Nunes. 2021. “The Right to Be Forgotten and COVID-19: Privacy versus Public Interest.” *Acta Bioethica* 27 (1): 59–67.

Dauda, Bege, and Kris Dierickx. 2013. “Benefit Sharing: An Exploration on the Contextual Discourse of a Changing Concept.” *BMC Medical Ethics* 14 (1): 36. https://doi.org/10.1186/1472-6939-14-36.

Devriendt, Thijs, Mahsa Shabani, and Pascal Borry. 2021. “Data Sharing in Biomedical Sciences: A Systematic Review of Incentives.” *Biopreservation and Biobanking* 19 (3): 219–27. https://doi.org/10.1089/bio.2020.0037.

EHDS2 Pilot. 2023. “Use Cases - EHDS2 Pilot - Official Website.” February 2, 2023. https://ehds2pilot.eu/use-cases/.

Erikainen, Sonja, Phoebe Friesen, Leah Rand, Karin Jongsma, Michael Dunn, Annie Sorbie, Matthew McCoy, et al. 2021. “Public Involvement in the Governance of Population-Level Biomedical Research: Unresolved Questions and Future Directions.” *Journal of Medical Ethics* 47 (7): 522–25. https://doi.org/10.1136/medethics-2020-106530.

European Commission. 2019. “European Data Strategy.” 2019. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.

European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS). 2022. “EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space.” July 12, 2022. https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-032022-proposal_en.

European Union. 2012. *Charter of Fundamental Rights of the European Union*. *OJ C*. Vol. 326. http://data.europa.eu/eli/treaty/char_2012/oj/eng.

Ferretti, Agata, and Effy Vayena. 2022. “In the Shadow of Privacy: Overlooked Ethical Concerns in COVID-19 Digital Epidemiology.” *Epidemics* 41 (December): 100652. https://doi.org/10.1016/j.epidem.2022.100652.

Findata. 2022. “Data Protection and the Processing of Personal Data.” Findata. March 14, 2022. https://findata.fi/en/about-findata/data-protection-and-the-processing-of-personal-data/.

Fontana, Gianluca, Saira Ghafur, Lydia Torne, Jonathan Goodman, and Ara Darzi. 2020. “Ensuring That the NHS Realises Fair Financial Value from Its Data.” *The Lancet Digital Health* 2 (1): e10–12. https://doi.org/10.1016/S2589-7500(19)30225-0.

Francis, Leslie, and John G Francis. 2017. “Data Re-Use and the Problem of Group Identity.” *Studies in Law, Politics, and Society* 73 (June): 141–64.

Ganguli-Mitra, Agomoni. 2012. “Benefit-Sharing, Biobanks and Vulnerable Populations.” In *Trust in Biobanking*, edited by Peter Dabrock, Jochen Taupitz, and Jens Ried, 257–66. Veröffentlichungen Des Instituts Für Deutsches, Europäisches Und Internationales Medizinrecht, Gesundheitsrecht Und Bioethik Der Universitäten Heidelberg Und Mannheim. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-78845-4_16.

Geneviève, Lester Darryl, Andrea Martani, Bernice Simone Elger, and Tenzin Wangmo. 2021. "Individual Notions of Fair Data Sharing from the Perspectives of Swiss Stakeholders." *BMC Health Services Research* 21 (1): 1007. https://doi.org/10.1186/s12913-021-06906-2.

Geneviève, Lester Darryl, Andrea Martani, Thomas Perneger, Tenzin Wangmo, and Bernice Simone Elger. 2021. "Systemic Fairness for Sharing Health Data: Perspectives From Swiss Stakeholders." *Frontiers in Public Health* 9. https://www.frontiersin.org/articles/10.3389/fpubh.2021.669463.

Golinelli, Davide, Fabrizio Toscano, Andrea Bucci, and Gherardo Carullo. 2018. "Transferring Health Big Data within the European Legal Framework: What Role for National Healthcare Services?" *Journal of Law and Medicine* 26 (2): 488–93.

Grewal, Norah, and Ainsley J. Newson. 2021. "The Perils of a Broad Approach to Public Interest in Health Data Research: A Response to Ballantyne and Schaefer." *Journal of Medical Ethics* 47 (8): 580–82. https://doi.org/10.1136/medethics-2020-106729.

Gross, Marielle S., Amelia J. Hood, Joshua C. Rubin, and Robert C. Miller. 2022. "Respect, Justice and Learning Are Limited When Patients Are Deidentified Data Subjects." *Learning Health Systems* 6 (3): e10303. https://doi.org/10.1002/lrh2.10303.

Haddow, Gillian, Graeme Laurie, Sarah Cunningham-Burley, and Kathryn G. Hunter. 2007. "Tackling Community Concerns about Commercialisation and Genetic Research: A Modest Interdisciplinary Proposal." *Social Science & Medicine* 64 (2): 272–82. https://doi.org/10.1016/j.socscimed.2006.08.028.

Hansen, Written Johan, Petra Wilson, Eline Verhoeven, Madelon Kroneman, Robert Verheij, and Evert-Ben van Veen. 2021. "Assessment of the EU Member States' Rules on Health Data in the Light of GDPR," 262.

Hepgul, Nilay, Katherine E. Sleeman, Alice M. Firth, Anna Johnston, James T. H. Teo, William Bernal, Richard J. B. Dobson, and Irene J. Higginson. 2019. "In Response to Ballantyne and Schaefer's

'Consent and the Ethical Duty to Participate in Health Data Research.'" *Journal of Medical Ethics* 45 (5): 351–52. https://doi.org/10.1136/medethics-2018-105271.

Involve, Carnegie UK Trust, and Understanding Patient Data. 2018. "Data for Public Benefit: Balancing the Risks and Benefits of Data Sharing." April 2018. https://understandingpatientdata.org.uk/data-public-benefit-risks-and-benefits-data-sharing.

Johnson, Tess, Konrad Kollnig, and Pierre Dewitte. 2022. "Towards Responsible, Lawful and Ethical Data Processing: Patient Data in the UK." *Internet Policy Review* 11 (1). https://policyreview.info/articles/analysis/towards-responsible-lawful-and-ethical-data-processing-patient-data-uk.

Jones, Kerina H., Graeme Laurie, Leslie Stevens, Christine Dobbs, David V. Ford, and Nathan Lea. 2017. "The Other Side of the Coin: Harm Due to the Non-Use of Health-Related Data." *International Journal of Medical Informatics* 97 (January): 43–51. https://doi.org/10.1016/j.ijmedinf.2016.09.010.

Kalkman, Shona, Johannes van Delden, Amitava Banerjee, Benoît Tyl, Menno Mostert, and Ghislaine van Thiel. 2022. "Patients' and Public Views and Attitudes towards the Sharing of Health Data for Research: A Narrative Review of the Empirical Evidence." *Journal of Medical Ethics* 48 (1): 3–13. https://doi.org/10.1136/medethics-2019-105651.

Kaplan, Bonnie. 2016. "How Should Health Data Be Used?: Privacy, Secondary Use, and Big Data Sales." *Cambridge Quarterly of Healthcare Ethics* 25 (2): 312–29. https://doi.org/10.1017/S0963180115000614.

Kuntsman, Adi, Esperanza Miyake, and Sam Martin. 2019. "Re-Thinking Digital Health: Data, Appisation and the (Im)Possibility of 'Opting Out.'" *DIGITAL HEALTH* 5 (January): 2055207619880671. https://doi.org/10.1177/2055207619880671.

Ledford, Heidi. 2019. "Millions of Black People Affected by Racial Bias in Health-Care Algorithms." *Nature* 574 (7780): 608–9. https://doi.org/10.1038/d41586-019-03228-6.

Lounsbury, Olivia, Lily Roberts, Jonathan R. Goodman, Philippa Batey, Lenny Naar, Kelsey M. Flott, Anna Lawrence-Jones, Saira Ghafur, Ara Darzi, and Ana Luisa Neves. 2021. "Opening a 'Can of Worms' to Explore the Public's Hopes and Fears About Health Care Data Sharing: Qualitative Study." *Journal of Medical Internet Research* 23 (2): e22744. https://doi.org/10.2196/22744.

Malgieri, Gianclaudio, and Jędrzej Niklas. 2020. "Vulnerable Data Subjects." *Computer Law & Security Review* 37 (July): 105415. https://doi.org/10.1016/j.clsr.2020.105415.

Marelli, Luca, Giuseppe Testa, and Ine Van Hoyweghen. 2021. "Big Tech Platforms in Health Research: Re-Purposing Big Data Governance in Light of the General Data Protection Regulation's Research Exemption." *Big Data & Society* 8 (1): 20539517211018783. https://doi.org/10.1177/20539517211018783.

McCarty, Catherine A., Rex L. Chisholm, Christopher G. Chute, Iftikhar J. Kullo, Gail P. Jarvik, Eric B. Larson, Rongling Li, et al. 2011. "The eMERGE Network: A Consortium of Biorepositories Linked to Electronic Medical Records Data for Conducting Genomic Studies." *BMC Medical Genomics* 4 (1): 13. https://doi.org/10.1186/1755-8794-4-13.

McCoy, Matthew S., Steven Joffe, and Ezekiel J. Emanuel. 2020. "Sharing Patient Data Without Exploiting Patients." *JAMA* 323 (6): 505–6. https://doi.org/10.1001/jama.2019.22354.

McLennan, Stuart, Leo Anthony Celi, and Alena Buyx. 2020. "COVID-19: Putting the General Data Protection Regulation to the Test." *JMIR Public Health and Surveillance* 6 (2): e19279. https://doi.org/10.2196/19279.

McMahon, Aisling, Alena Buyx, and Barbara Prainsack. 2020. "Big Data Governance Needs More Collective Responsibility: The Role of Harm Mitigation in the Governance of Data Use in Medicine and Beyond." *Medical Law Review* 28 (1): 155–82. https://doi.org/10.1093/medlaw/fwz016.

Meszaros, Janos, and Chih-hsing Ho. 2019. “Building Trust and Transparency? Challenges of the Opt-out System and the Secondary Use of Health Data in England.” *Medical Law International* 19 (2–3): 159–81. https://doi.org/10.1177/0968533219879975.

Müller, Sebastian. 2022. “Is There a Civic Duty to Support Medical AI Development by Sharing Electronic Health Records?” *BMC Medical Ethics* 23 (1): 134. https://doi.org/10.1186/s12910-022-00871-z.

Munung, Nchangwi Syntia, and Jantina de Vries. 2020. “Benefit Sharing for Human Genomics Research: Awareness and Expectations of Genomics Researchers in Sub-Saharan Africa.” *Ethics & Human Research* 42 (6): 14–20. https://doi.org/10.1002/eahr.500069.

Murtagh, Madeleine J., Mavis Machirori, Clara L. Gaff, Mwenza T. Blell, Jantina de Vries, Megan Doerr, Edward S. Dove, et al. 2021. “Engaged Genomic Science Produces Better and Fairer Outcomes: An Engagement Framework for Engaging and Involving Participants, Patients and Publics in Genomics Research and Healthcare Implementation.” *Wellcome Open Research* 6 (November): 311. https://doi.org/10.12688/wellcomeopenres.17233.1.

NHS England. 2023. “National Data Opt-Out.” NHS Digital. July 5, 2023. https://digital.nhs.uk/services/national-data-opt-out.

O’Brien, Emily C., Ana Maria Rodriguez, Hye-Chung Kum, Laura E. Schanberg, Marcy Fitz-Randolph, Sean M. O’Brien, and Soko Setoguchi. 2019. “Patient Perspectives on the Linkage of Health Data for Research: Insights from an Online Patient Community Questionnaire.” *International Journal of Medical Informatics* 127 (July): 9–17. https://doi.org/10.1016/j.ijmedinf.2019.04.003.

O’Doherty, Kieran C., Emily Christofides, Jeffery Yen, Heidi Beate Bentzen, Wylie Burke, Nina Hallowell, Barbara A. Koenig, and Donald J. Willison. 2016. “If You Build It, They Will Come: Unintended Future Uses of Organised Health Data Collections.” *BMC Medical Ethics* 17 (1): 54. https://doi.org/10.1186/s12910-016-0137-x.

Peters, Micah D. J., Christina M. Godfrey, Hanan Khalil, Patricia McInerney, Deborah Parker, and Cassia Baldini Soares. 2015. "Guidance for Conducting Systematic Scoping Reviews." *International Journal of Evidence-Based Healthcare* 13 (3): 141–46. https://doi.org/10.1097/XEB.0000000000000050.

Ploug, Thomas, and Søren Holm. 2017. "Informed Consent and Registry-Based Research - the Case of the Danish Circumcision Registry." *BMC Medical Ethics* 18 (1): 53. https://doi.org/10.1186/s12910-017-0212-y.

Porsdam Mann, Sebastian, Julian Savulescu, and Barbara J. Sahakian. 2016. "Facilitating the Ethical Use of Health Data for the Benefit of Society: Electronic Health Records, Consent and the Duty of Easy Rescue." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2083): 20160130. https://doi.org/10.1098/rsta.2016.0130.

Prainsack, Barbara. 2018. "The 'We' in the 'Me': Solidarity and Health Care in the Era of Personalized Medicine." *Science, Technology, & Human Values* 43 (1): 21–44.

Prainsack, Barbara, Seliem El-Sayed, Nikolaus Forgó, Łukasz Szoszkiewicz, and Philipp Baumer. 2022. "Data Solidarity: A Blueprint for Governing Health Futures." *The Lancet Digital Health* 4 (11): e773–74. https://doi.org/10.1016/S2589-7500(22)00189-3.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space. 2022. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197&qid=1682428844692.

Ramsay, Michèle. 2022. "African Genomic Data Sharing and the Struggle for Equitable Benefit." *Patterns* 3 (1): 100412. https://doi.org/10.1016/j.patter.2021.100412.

Ravinetto, Raffaella, Pier Davide Guenzi, Pierre Massat, and Gianluca Gaidano. 2014. "Globalisation of Clinical Trials and Ethics of Benefit Sharing." *The Lancet Haematology* 1 (2): e54–56. https://doi.org/10.1016/S2352-3026(14)00004-0.

Ravinetto, Raffaella, and Jerome Amir Singh. 2022. “Responsible Dissemination of Health and Medical Research: Some Guidance Points.” *BMJ Evidence-Based Medicine*, September. https://doi.org/10.1136/bmjebm-2022-111967.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance). 2016. *OJ L*. Vol. 119. http://data.europa.eu/eli/reg/2016/679/oj/eng.

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA Relevance). 2022. *OJ L*. Vol. 152. http://data.europa.eu/eli/reg/2022/868/oj/eng.

Richter, Gesine, Christoph Borzikowsky, Bimba Franziska Hoyer, Matthias Laudes, and Michael Krawczak. 2021. “Secondary Research Use of Personal Medical Data: Patient Attitudes towards Data Donation.” *BMC Medical Ethics* 22 (December): 164. https://doi.org/10.1186/s12910-021-00728-x.

Rocher, Luc, Julien M. Hendrickx, and Yves-Alexandre de Montjoye. 2019. “Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models.” *Nature Communications* 10 (1): 3069. https://doi.org/10.1038/s41467-019-10933-3.

Ross, Michael W., Martin Y. Iguchi, and Sangeeta Panicker. 2018. “Ethical Aspects of Data Sharing and Research Participant Protections.” *The American Psychologist* 73 (2): 138–45. https://doi.org/10.1037/amp0000240.

Sabatello, Maya, Daphne O. Martschenko, Mildred K. Cho, and Kyle B. Brothers. 2022. “Data Sharing and Community-Engaged Research.” *Science* 378 (6616): 141–43. https://doi.org/10.1126/science.abq6851.

Schinas, Margaritis. 2023. “Schinas, Margaritis (Vice-President of the European Commission for Promoting Our European Way of Life). Event on the European Health Data Space, European Parliament, Brussels. Opening Remarks.” February 28.

Seastedt, Kenneth P., Patrick Schwab, Zach O’Brien, Edith Wakida, Karen Herrera, Portia Grace F. Marcelo, Louis Agha-Mir-Salim, et al. 2022. “Global Healthcare Fairness: We Should Be Sharing More, Not Less, Data.” *PLOS Digital Health* 1 (10): e0000102. https://doi.org/10.1371/journal.pdig.0000102.

Shabani, Mahsa. 2022. “Will the European Health Data Space Change Data Sharing Rules?” *Science* 375 (6587): 1357–59. https://doi.org/10.1126/science.abn4874.

Shabani, Mahsa, and Sami Yilmaz. 2022. “Lawfulness in Secondary Use of Health Data: Interplay between Three Regulatory Frameworks of GDPR, DGA & EHDS.” *Technology and Regulation* 2022 (December): 128–34. https://doi.org/10.26116/techreg.2022.013.

Sharon, Tamar. 2018. “When Digital Health Meets Digital Capitalism, How Many Common Goods Are at Stake?” *Big Data & Society* 5 (2): 2053951718819032. https://doi.org/10.1177/2053951718819032.

Slavkovik, Marija. 2023. “Mythical Ethical Principles for AI and How to Attain Them.” In *Human-Centered Artificial Intelligence: Advanced Lectures*, edited by Mohamed Chetouani, Virginia Dignum, Paul Lukowicz, and Carles Sierra, 275–303. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-24349-3_15.

Slokenberga, Santa. 2020. “Setting the Foundations: Individual Rights, Public Interest, Scientific Research and Biobanking.” *GDPR and Biobanking* 43 (May): 11–30. https://doi.org/10.1007/978-3-030-49388-2_2.

Street, Jackie, Belinda Fabrianesi, Carolyn Adams, Felicity Flack, Merran Smith, Stacy M. Carter, Sean Lybrand, et al. 2021. “Sharing Administrative Health Data with Private Industry: A Report on Two Citizens’ Juries.” *Health Expectations* 24 (4): 1337–48. https://doi.org/10.1111/hex.13268.

Sweeney, Latanya. 2015. “Only You, Your Doctor, and Many Others May Know.” *Technology Science*, September. https://techscience.org/a/2015092903/.

Tangcharoensathien, Viroj, Jirawan Boonperm, and Pongpisut Jongudomsuk. 2010. “Sharing Health Data: Developing Country Perspectives.” *Bulletin of the World Health Organization* 88 (6): 468–69. https://doi.org/10.2471/BLT.10.079129.

Taylor, Linnet. 2017. “What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally.” *Big Data & Society* 4 (2): 2053951717736335. https://doi.org/10.1177/2053951717736335.

Taylor, Mark J., and Tess Whitton. 2020. “Public Interest, Health Research and Data Protection Law: Establishing a Legitimate Trade-Off between Individual Control and Research Access to Health Data.” *Laws* 9 (1): 6. https://doi.org/10.3390/laws9010006.

TEHDAS (Joint Action Towards the European Health Data Space). 2022a. “Report on Secondary Use of Health Data through European Case Studies.” February 28, 2022. https://tehdas.eu/results/tehdas-suggests-options-to-overcome-data-barriers/.

———. 2022b. “TEHDAS Consultation: People Support Health Data Use with Solid Safeguards.” Tehdas. June 30, 2022. https://tehdas.eu/results/tehdas-consultation-people-support-health-data-use-with-solid-safeguards/.

———. 2023a. “Lessons Learnt and Recommendations to Foster GDPR Compliant Data Altruism Mechanisms for the EHDS (Unpublished).”

———. 2023b. “Qualitative Study to Assess Citizens’ Perception of Sharing Health Data for Secondary Use and Recommendations on How to Engage Citizens in the EHDS.” March 31, 2023. https://tehdas.eu/results/tehdas-consultation-citizens-support-the-secondary-use-of-health-data-when-it-matches-their-ethical-values/.

Terzis, Petros. 2020. “Onward for the Freedom of Others: Marching beyond the AI Ethics.” In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 220–29. Barcelona Spain: ACM. https://doi.org/10.1145/3351095.3373152.

———. 2022. “Compromises and Asymmetries in the European Health Data Space.” *European Journal of Health Law* 1 (aop): 1–19. https://doi.org/10.1163/15718093-bja10099.

The Wellcome Trust. 2017. “The One-Way Mirror: Public Attitudes to Commercial Access to Health Data,” November. https://doi.org/10.6084/m9.figshare.5616448.v1.

Thomas, Aliki, Stuart Lubarsky, Steven J. Durning, and Meredith E. Young. 2017. “Knowledge Syntheses in Medical Education: Demystifying Scoping Reviews.” *Academic Medicine* 92 (2): 161. https://doi.org/10.1097/ACM.0000000000001452.

Townend, David. 2018. “Conclusion: Harmonisation in Genomic and Health Data Sharing for Research: An Impossible Dream?” *Human Genetics* 137 (8): 657–64. https://doi.org/10.1007/s00439-018-1924-x.

Tully, Mary P., Lamiece Hassan, Malcolm Oswald, and John Ainsworth. 2019. “Commercial Use of Health Data—A Public ‘Trial’ by Citizens’ Jury.” *Learning Health Systems* 3 (4): e10200. https://doi.org/10.1002/lrh2.10200.

Vayena, Effy, Joan Dzenowagis, John S Brownstein, and Aziz Sheikh. 2018. “Policy Implications of Big Data in the Health Sector.” *Bulletin of the World Health Organization* 96 (1): 66–68. https://doi.org/10.2471/BLT.17.197426.

Waind, Elizabeth. 2020. “Trust, Security and Public Interest: Striking the Balance A Narrative Review of Previous Literature on Public Attitudes towards the Sharing, Linking and Use of Administrative Data for Research.” *International Journal of Population Data Science* 5 (3): 1368. https://doi.org/10.23889/ijpds.v5i3.1368.

Williams, Gemma A, and Nick Fahy. 2019. “BUILDING AND MAINTAINING PUBLIC TRUST TO SUPPORT THE SECONDARY USE OF PERSONAL HEALTH DATA.”

Winkler, Eva C, Martin Jungkunz, Adrian Thorogood, Vincent Lotz, and Christoph Schickhardt. 2023. “Patient Data for Commercial Companies? An Ethical Framework for Sharing Patients’ Data with for-Profit Companies for Research.” *Journal of Medical Ethics*, May, jme-2022-108781. https://doi.org/10.1136/jme-2022-108781.

Zou, James, and Londa Schiebinger. 2018. “AI Can Be Sexist and Racist — It’s Time to Make It Fair.” *Nature* 559 (7714): 324–26. https://doi.org/10.1038/d41586-018-05707-8.