E&HR

The Ethical Implications of Big Data Research in Public Health:
"Big Data Ethics by Design" in the UK-REACH Study

Ruby Reed-Berendt, Edward S. Dove, and Manish Pareek, on behalf of the UK-REACH Study Collaborative Group

ABSTRACT In this article, we analyze legal and ethical issues raised in Big Data health research projects in the Covid-19 era and consider how these issues might be addressed in ways that advance positive values (e.g., furtherance of respect for persons and accordance with relevant legal frameworks) while mitigating or eliminating any negative aspects (e.g., exacerbation of social inequality and injustice). We apply this analysis specifically to UK-REACH (The United Kingdom Research Study into Ethnicity and Covid-19 Outcomes in Healthcare Workers), a project with which we are involved. We argue that Big Data projects like UK-REACH can be conducted in an ethically robust manner and that funders and sponsors ought to encourage similar projects to drive better evidence-based public policy in public health. As part of this, we advocate that a *Big Data ethics-by-design* approach be undertaken when such projects are constructed. This principle extends the work of those who advocate ethics by design by addressing prominent issues in Big Data health research projects; it holds that ethical values and principles in Big Data health research projects are best adhered to when they are already integrated into the project aims and methods at the design stage. In advocating this principle, we present a unique perspective regarding pressing ethical problems around large-scale, data-driven Covid-19 research, as well as legal issues associated with processing ostensibly anonymized health data.

KEYWORDS human research ethics, Big Data, Big Data health research, Big Data ethics, Covid-19, Covid-19 research, UK-REACH study, health data, data linkage

Reed-Berendt, R., E. S. Dove, and M. Pareek, on behalf of the UK-REACH Study Collaborative Group, "The Ethical Implications of Big Data Research in Public Health: 'Big Data Ethics by Design' in the UK-REACH Study," *Ethics & Human Research* 44, no. 1 (2022): 2-17. DOI: 10.1002/eahr.500111

Research projects involving large-scale processing of health data and data linkage (so-called Big Data projects), were already well underway before the Covid-19 pandemic struck in March 2020.[1] But without question, since the pandemic has spread across the globe, projects of this nature have accelerated and received heightened attention for the benefits they can bring to science and medicine, as well as public policy in the public health context.[2]

These projects are not, however, immune from conceptual and practical challenges at both the design and implementation stages, including those of a legal and ethical nature. In all circumstances, using patient data and other forms of health data in research—and linking those data to other categories of data such as employment and administrative data—raises questions around, inter alia, consent, privacy, confidentiality, justice, and respect for human rights. Such questions become even more pronounced in a pandemic that has affected all of our daily lives and had a disproportionate impact on certain groups (for example, ethnic minorities).[3] In this context, additional layers of vulnerability arise—not just because of the potentially vulnerable status of some of the participants (or "data subjects") who may have suf-

fered the ill effects of the disease but also because of other existing vulnerabilities, including socioeconomic status, age, gender, religion, and ethnic identity. Such vulnerabilities may be exacerbated in the lifespan of the project, either directly or indirectly in response to the nature of the investigation.

These concerns exist whether the populations under study are entire swaths of the general public or more targeted groups such as health care workers (HCWs), older adults residing in care homes, transport workers, or people living with disabilities. Thus, in these Big Data health research projects in the Covid-19 era, careful attention ought to be paid to (a) identifying the specific legal and ethical issues raised by the project's aims and methods and (b) addressing those identified issues in ways that advance positive aspects (e.g., furtherance of justice and respect for persons and accordance with the letter and spirit of the relevant legal frameworks) while mitigating or eliminating any negative aspects (e.g., risk of stigmatization, the violation of law, and the exacerbation of social inequality and injustice).

In this article, we apply this two-part examination to the UK-REACH study (The United Kingdom Research Study into Ethnicity and Covid-19 Outcomes in Healthcare Workers),[4] with which we are involved (ESD as coinvestigator, RRB as research associate, and MP as principal investigator; a full list of the UK-REACH Collaborative Group members appears in the online appendix; see the "Supporting Information" section below).[5] Collaborating with HCW regulatory bodies and professional bodies or associations, over a period of at least 12 months, UK-REACH is a mixed-methods study aiming to investigate if, how, and why ethnicity affects Covid-19 clinical outcomes in HCWs across the United Kingdom. The project is providing evidence through interlinked work packages (WPs), which investigate instances of Covid-19 among HCWs, as well as their experiences of working during the pandemic. The project also encompasses a multiprofessional, national stakeholder group, including the HCW statutory regulators, Royal Colleges, ethnic minority professionals' associations, and ethnic minority HCWs, and is facilitating rapid dissemination and translation of the research findings for HCWs, employers, and policy-makers (see table 1 for a brief summary of each WP).[6]

For the purposes of this article, we discuss the research of WP1, which is undertaking an expedited linkage and analysis of anonymized HCW registration, employment, and health care datasets, within a trusted research environment (the SAIL Databank at Swansea University in Wales),[7] to calculate the incidence of, and outcomes from, Covid-19 among HCWs. Linking such datasets (employment, professional registration, and health data) in the Covid-19 context may be novel and therefore may present different sets of challenges, meaning it is necessary to consider what ethical issues arise out of this activity. Equally, research on health care workers in the context of a pandemic in and of itself raises ethical concerns,[8] with sensitivity heightened through UK-REACH's focus on ethnicity. In what follows, we identify the legal and ethical issues arising through WP1 and propose how this work can be done in a manner that is ethically, legally, and socially acceptable. We believe that our framework of analysis has value for a wide array of Big Data projects both underway and being planned, including those that may not be directly related to Covid-19 research.

> Beyond the work of articulating and interrogating the benefits of Big Data research in public health, consideration should be given to how to balance them with the interests of privacy and justice.

To ensure broad purchase, we focus on unpacking the ethical implications of using Big Data (for us, meaning data processed with significant volume, variety, and velocity)[9] in public health research, specifically exploring these questions: What ethical concerns arise from conducting public health research, using Big Data on Covid-19 outcomes, HWCs, and ethnicity? How can public health research involving Big Data be conducted in a way that is ethically acceptable? And what value can an ethical approach bring to using Big Data research

projects in public health alongside the existing legal framework?

Through applying the two-part examination and exploring each of the above questions, we argue that Big Data projects like UK-REACH can be conducted in an ethically robust manner, and indeed similar projects ought to be encouraged by funders and sponsors to drive better evidence-based public policy in public health. However, we further argue that it is crucial that a *Big Data ethics-by-design* approach be undertaken when constructing such projects. This principle extends the work by scholars who have advocated an ethics by design approach, which has become increasingly prominent in the field of artificial intelligence (AI) and data ethics.[10] Our proposed Big Data ethics-by-design approach holds that ethics in Big Data projects are best adhered to when they are already integrated into the project aims and methods at the design stage. This should reflect an engaged, transparent, and reflexive approach that demonstrates what ongoing measures are being taken to protect and promote participants' data and interests, as well as to enable efficient and effective responses to any concerns that are raised. Moreover, the approach will allow any new challenges to be rapidly and properly considered to ensure harms are identified and addressed where required. In advocating Big Data ethics by design, our article brings a unique perspective regarding some of the pressing ethical problems around large-scale, data-driven Covid-19 research, as well as the legal issues associated with processing ostensibly anonymized health data.

We begin with an analysis of the relevant legal framework, before turning to analysis of relevant ethical considerations. We conclude, drawing on learning from the UK-REACH project, by advocating a Big Data ethics-by-design approach, whereby public health research projects using Big Data, including and beyond UK-REACH, can undertake research in a manner that is ethically acceptable.

THE LEGAL FRAMEWORK OF UK-REACH

Any Big Data project involving the processing of "personal data," which, according to article 4(1) of the European Union General Data Protection Regulation (GDPR) 2016/679, means any information relating to an identified or identifiable natural person (i.e., human individual), must ensure that it is meeting its legal obligations with respect to that data.[11] The contours of these obligations will necessarily differ for each project dependent on its jurisdiction and the data being processed, and each project should seek to identify and consider its relevant legal framework. We highlight

Table 1.
Overview of Each Work Package in UK-REACH

Work package	*Overview*
WP1	Linkage and analysis of anonymized data on health care workers (HCWs), including National Health Service datasets, human resources data, professional registration data, and Covid-19 datasets to analyze incidences and outcomes of Covid-19 across HCWs
WP2	Longitudinal cohort study of approximately 18,000 HCWs through baseline and follow-up questionnaires on their experience of working during the Covid-19 pandemic
WP3	Exploration of legal and ethical issues raised by the project, through desk-based policy analysis and qualitative semistructured interviews with around 20 key opinion leaders in health and health research (interviews by ED and RRB)
WP4	Qualitative study examining HCW's knowledge, behavior, and perceptions of risk in relation to Covid-19 through interviews and focus groups with approximately 150 HCWs
WP5	A multiprofessional national stakeholder group of HCWs, professional organizations, and regulatory bodies designed to inform how the study is conducted and to facilitate the dissemination and translation of findings into policy
WP6	An immunology study focused on understanding whether differences exist in the scale, profile, and duration of the immune response in HCWs from diverse ethnic backgrounds

the three key frameworks that must be considered by UK-REACH as part of its compliance obligation, two of which are situated in the common law, and the third of which is situated in statutory law.

The common law: confidentiality and private information. The common law duty of confidentiality holds that where an individual has a reasonable expectation of privacy with respect to information given in confidence and where a person receiving that information knows or ought to know that the other can reasonably expect their privacy to be protected, such information should be kept confidential and not disclosed except under certain conditions.[12] The duty may arise through professional or contractual relationships, as well as where information is imparted in circumstances importing an obligation of confidence (e.g., research involving human participants in which information of a confidential nature is disclosed to researchers). The obligation is not to pass on the information to any third parties without justification and not to make the information (publicly) available in identifiable form.

In considering how the duty of confidentiality is applicable in Big Data research, close attention should be paid to the sources of data. For UK-REACH, it is relevant through use of both health care and employment data. It is well-established in case law[13] that health care information is subject to the duty of confidentiality; this is also codified in codes of conduct from professional regulators.[14] The contractual employment relationship also requires an employer (e.g., the National Health Service [NHS]) to keep employees' data confidential. Therefore, the data subjects within WP1 have a reasonable expectation that their information, be it for health or employment, will remain confidential and not be used in ways that deviate from a reasonable expectation of privacy, unless there is a specific legal basis that lifts the duty of confidentiality.

Another part of the common law that now applies to use of information of this kind is tort law. In the United Kingdom, over the past few years, a specific tort of "misuse of private information" has been crafted that protects individuals against the wrongful use of their private information.[15] The scope of this tort is wider than the common law duty of confidentiality, which applies only to confidential information (that invariably arises in the doctor-patient relationship and covers almost all forms of patient information), which need not be of a personal nature. Private information, by contrast, applies to information that is personal and by its very nature is considered "private" (such as employment data, medical files, and hospital records). Whereas in a claim for breach of confidentiality, a court historically has looked to a consideration of whether (a) the information has the necessary quality of confidence, (b) the information was imparted in circumstances importing an obligation of confidence, and (c) there was an unauthorized use of the information causing detriment,[16] in a claim for misuse of private information, a court will consider whether (a) the information in question is information over which the claimant had a reasonable expectation of privacy and (b), if so, whether the claimant's right to privacy under article 8 of the European Convention on Human Rights (ECHR)[17] outweighs the defendant's article 10 ECHR right to freedom of expression.[18] In practice, and despite these legal nuances, Big Data projects involving the processing of personal data—especially if the data are gathered in a health-related context—must ensure that they are meeting their legal obligations across both of these common law domains.

U.K. data protection law. In addition to duties under the common law, the statutory framework for the processing of personal data by UK-REACH is provided by the GDPR (which, following Brexit, has become the U.K. GDPR) and its transposition into national law and supplement provided by the U.K.'s Data Protection Act 2018. To process lawfully any "personal data," as defined in article 4(1) of the GDPR, one must have a legal basis as stipulated in article 6(1) of the GDPR. This article 6(1) legal basis should be read in conjunction with article 9(1) and article 9(2), which require "special category" personal data (including data concerning ethnicity and data concerning health) to meet one of ten lawful exemptions from the prohibition against processing such data.

The GDPR is clear that the principles of data protection apply to *all* information that concerns identified or identifiable persons. This includes information that has undergone "pseudonymization"[19] but can still be identified through the use of additional information (e.g., encoding of a dataset that can be connected to a specific individual with a code key).[20] (We note that pseudonymization and anonymization are GDPR-related con-

cepts, both of which are distinct from "deidentification," which is a term more commonly used in data privacy statutes in the United States and elsewhere and refers to the removal of many, but not necessarily all, identifiers in a given dataset.) However, the GDPR will not apply to the processing of data that are anonymized per the standard of recital 26, i.e., information that does not relate to an identified or identifiable person or personal data rendered anonymous in such a manner that the data subject is not or is no longer identifiable (for example, through rendering data down to an aggregated level or converting data into statistics so that individuals can no longer be identified from them).[21] Importantly, whether an individual data item can be considered anonymous or not requires case-by-case evaluation. Collected material can contain detailed information on individuals (e.g., rare diseases, postcode and occupation, or a sufficient amount of data of different types) that makes them indirectly identifiable.

As UK-REACH is collecting and processing data relating to identified persons, the GDPR is applicable. The data used also will include special categories of data within the meaning of article 9 of the GDPR, namely, data relating to ethnicity and data concerning health. Although the GDPR is no longer applicable once data are adequately anonymized, it is applicable to the processing of those data for the purposes of achieving anonymization. In other words, when data are initially gathered and before they have been anonymized, they are subject to the full requirements of data protection law. There is, therefore, a need to stipulate the legal basis for processing the data in WP1 prior to the removal of direct identifiers or further processing of data following deidentification that does not meet the Recital 26 standard (this will be considered further from an ethical perspective below).

For UK-REACH, the lawful basis for data processing is article 6(1)(e) of the GDPR and the special category exception to process special-category data is article 9(2)(i) of the GDPR. Articles 6(1)(e) and 9(2)(i) require, respectively, that the data processing is necessary for the performance of a task carried out in the public interest and that it is necessary for reasons of public interest in the area of public health. Therefore, UK-REACH (and any project that seeks to rely on these articles to process data) needs to demonstrate that its data processing activities are necessary and in the public interest. The question how we might demonstrate that the project is in the public interest will be considered further below. We also note that other lawful bases are available to projects such as UK-REACH, for instance, article 9(2)(a) (concerning explicit consent) or 9(2)(j) (concerning processing for the purpose of scientific research). As with public interest, each of these provisions come with their own confines; for example, consent requires researchers to ensure they use only the data in line with the specific consent (and is not relevant for WP1, in which consent has not been sought), and those relying on scientific research must ensure it is based on domestic law that is proportionate to the aim pursued and must provide for suitable and specific measures to safeguard the fundamental rights and interests of the data subjects.

Human rights law. Finally, Big Data research projects conducted in the United Kingdom (and elsewhere in Europe) must give due consideration to the relevant human rights protected under the ECHR. These rights form part of U.K. law under the Human Rights Act 1998. Relevant rights in the Human Rights Act are those in articles 8 and 14. Article 8, which addresses the right to respect for privacy and family life, is engaged because the UK-REACH project involves the collection, storage, and use of detailed data relating to broad aspects of the private lives of data subjects.[22] The collection, storage, and use of data[23] must therefore be compatible with respect for the privacy of that information, with any interference justified as necessary and proportionate to the legitimate interest pursued.[24]

Article 14, which holds that all the rights within the ECHR must be guaranteed without discrimination, is particularly important given UK-REACH's focus on race and ethnicity, two of its identified characteristics. The analytical focus on ethnicity creates a potential risk of differential treatment or disproportionate negative impact on a particular group. Because of the health and employment context, this impact could affect individuals' livelihoods and/or health status—both of which engage article 8. UK-REACH should be mindful of any direct or indirect discrimination arising through the project itself, or how its findings are used to make policy. Even policies that appear neutral, for example, Covid-19 physical-distancing restrictions or work re-

strictions (which at first glance apply to everyone in the same way) may in fact have a worse effect on some people than others, such as members of ethnic minority communities. Any potential negative impact should be identified and mitigated.

The above provides a summary of the three constituent elements of UK-REACH's legal framework. As we have noted, the legal questions arising will be specific to the project in question and its jurisdiction. We advocate for any project to carefully consider at the outset which legal frameworks are relevant and take appropriate steps to ensure its compliance. Having considered this, we now turn to analyze the relevant ethical considerations for UK-REACH and Big Data projects.

ETHICS OF BIG DATA AND PUBLIC HEALTH RESEARCH

Regardless of the legal particulars of a research project, even where it meets its legal obligations, it does not necessarily follow that its use of data will be ethically acceptable, nor does it mean that no obligations are owed in respect of its data subjects. There remains a need for use of data in research to be ethically and publicly acceptable. To help projects consider how to ensure such use, Xafis and colleagues have proposed an ethical decision-making framework[25] (which we draw on further below). In line with their core characteristics, we consider that UK-REACH meets the definition of a "Big Data" project as follows:

- Volume: Detailed data relating to in excess of 1 million people are being gathered.
- Variety: Data come from diverse sources and sectors, including not just health data but also employment and professional registration data.
- Velocity: Expedited linkage and analysis of multiple datasets are being carried out, facilitated by the rapid sharing of data across multiple organizations.

As outlined by the Nuffield Council on Bioethics in their 2015 report on data-driven health research, the use and linkage of Big Data bring significant opportunities for research,[26] such as the advancement of medical treatment or, in the case of UK-REACH, policy outputs to benefit HCWs. But data usage in research is not without risk of harm to individuals or their interests.[27] Thus, it remains important to ensure that morally relevant interests are balanced with the benefits of using Big Data in research, in large part because the individual's consent is *not* being sought (unlike in the usual process for participation in research), meaning they have limited opportunity to control how their data are used.

However, ethical questions relating to Big Data are not the only consideration for UK-REACH. As an urgent public health research project,[28] it aims to inform public health responses to the pandemic. *Public health ethics* are therefore also relevant. Public health ethics ought to be distinguished from research ethics or clinical ethics in that, among other distinguishing features, the focus of moral conduct and pursuit of the good is less on any given individual and more on the collective or public good. Thus, ethical principles drawn from biomedical ethics, such as the principlist framework from Beauchamp and Childress, which emphasizes respect for autonomy, nonmaleficence, beneficence, and justice, arguably has less relevance than principles emphasizing, inter alia, solidarity,[29] reciprocity, social justice, collective well-being, and harm minimization.[30]

Therefore, we consider that projects like UK-REACH are situated at the intersection between the concerns of public health ethics and ethical use of Big Data. This requires a reexamination not only of how the key principles are conceptualized but also which is given the most weight (a point we discuss further below).

One key ethical concern we highlight for both Big Data and public health is to tackle vulnerabilities.[31] To explore the relationship between vulnerability, Big Data, and public health, we adopt the taxonomy proposed by Mackenzie, Rogers, and Dodds.[32] On this understanding, vulnerability can be inherent to the human condition, but it can also be situational; it is created or worsened by a person's individual context, be it social, political, or economic. Vulnerability also can be pathogenic in that it can be generated from responses aimed at ameliorating vulnerability or can arise from morally dysfunctional relationships or systems (i.e., those characterized by prejudice, oppression, and injustice).

All three of these types of vulnerability can feature in public health measures and Big Data research. For example, all data subjects in Big Data research projects, including those in public health that make use of their personal data, are inherently vulnerable to breaches of privacy; in general, the more data that are collected and used and the more "sensitive" the data are, the higher

the risk of privacy breaches and informational harms becomes. However, the unequal distribution of power between those who are the subject of the data and those who control them (the "Big Data divide")[33] means that data subjects with lower socioeconomic status tend to be less able to control what happens with their data and therefore are more situationally vulnerable and exposed to their misuse. Equally, pathogenic vulnerabilities may arise where groups of individuals (e.g., women or ethnic minorities) are excluded from the benefits of research or where group harms occur as an unintended consequence of Big Data research (e.g., research findings that lead, directly or indirectly, to discriminatory policies or stigmas against community groups).[34]

Similarly, in the public health context, although everyone may be inherently vulnerable to the ill effects of disease, the burden of disease falls more on certain populations who are situationally vulnerable. For example, the prevalence of obesity differs significantly based on socioeconomic status, gender, and ethnicity,[35] meaning that certain groups are situationally vulnerable to health conditions or mortality associated with obesity. It is also very important that public health responses be mindful of targeting specific groups in a manner that may engender further stigmatization of such groups and the generation of pathogenic vulnerability.

To better mitigate any negative impact of Big Data research in public health, it is crucial that those designing and managing such research identify and understand, as an overriding ethical concern, how vulnerabilities may be generated through the research activities.[36] We now proceed to explore these issues—and possible mitigation strategies—in more detail.

Ethical concerns for Big Data-driven public health research. Beyond the general public health ethical concerns noted above, we highlight two key reasons that an ethical approach is particularly important in the context of Big Data-driven public health research, especially during the Covid-19 pandemic: first, the impact of the pandemic on research and acceptability and, second, specific vulnerabilities that arise in this context.

- **Research during a pandemic.** Research that focuses on, and is undertaken during, a pandemic like Covid-19 faces a unique set of challenges. As we witnessed over the course of 2020 and 2021, pandemics bring significant change and uncertainty, where individuals are faced with new and shifting laws, policies, and regulations that restrict movement and other behaviors and that ebb and flow with respect to scope and magnitude depending on political choices and changes in the severity and spread of the virus (among other factors). This situation causes uncertainty for individuals on numerous levels, including whether their behavior and activities are legally (or morally) permissible at a given time, their ongoing financial security, or indeed what should happen to them if they catch the virus.

The uncertainty induced by the pandemic is likely to influence individuals' perception of risk, as well as the trust they place in law, policy, and regulation (especially when the rules appear to be subject to constant change, and not always in response to the science and epidemiological data).[37] This may mean that individual views about what constitutes acceptable use of their data, the level of concern about use of the data, or the trust they place in law to protect their data, may differ from what such views were in the prepandemic period (an issue that will be teased out in both WPs 3 and 4 in UK-REACH). The potential for a more skeptical view on data usage presents a challenge to the public acceptability and trustworthiness of research such as UK-REACH. An ethical approach offers the chance to assuage such doubts and support the public acceptability of research.

Moreover, the fast-paced legal and regulatory environment is also likely to bring novel challenges around what constitutes appropriate use of individual data. For example, contact-tracing measures, such as the recording of clients' contact details at hospitality establishments or the tracking of individuals' movements via mobile applications, raise new concerns around data protection and the extent of privacy rights and state surveillance measures in the context of a public health emergency.[38] The changing landscape raises the possibility that research on Covid-19 will need to consider novel kinds of data use and sharing that have not been previously addressed or even identified (so-called known unknowns and unknown unknowns, respectively). In such novel situations, it is not sufficient to merely ensure legal compliance; it is essential that there is ongoing ethical acceptability of the research, which means that the panoply of ethical issues arising from such data usages are appropriately identified and responded to in ways that respect participants. This includes respect

for the wider communities in which the research and its participants are situated (communities, we note, that may be personal and familial as well as professional).[39]

• **Vulnerability and Covid-19.** As discussed above, vulnerability is of key ethical concern for both Big Data research and public health measures; indeed, concerns relating to health inequalities have received renewed attention during the pandemic, with research indicating that most of the increased risk of infection and death from Covid-19 among people from ethnic minorities is explained by factors such as occupation, where people live, their household composition, and preexisting health conditions.[40] Research related to a pandemic should, as a general principle, also seek to consider what new situational vulnerabilities that arise out of the pandemic context have relevance to the project.

In the case of UK-REACH, this requires giving due consideration to any vulnerabilities related to ethnicity, socioeconomic status, and job role (HCW, for example).[41] One such situational vulnerability is an ostensibly heightened risk within these groups of contracting and suffering the ill effects of the virus,[42] but there are others not directly related to risk of the virus itself. For example, HCWs working in an intensive care unit are, because of their job, situationally vulnerable to mental distress, which arises from treating multiple Covid-19 patients in this environment.[43] These vulnerabilities are of further concern because they affect groups that already experience disadvantages from social inequality and discrimination; new situational vulnerabilities may layer upon existing pathogenic vulnerabilities in a way that further disadvantages such groups. Moreover, individuals at the intersection of these groups, such as ethnic minority HCWs, might experience the culmination of these various layers, opening them up to further vulnerability.

Although public health research focusing on ethnic minorities will often aim to address relevant vulnerabilities (indeed, this is an opportunity with projects like UK-REACH), it is also important to consider the potential for new pathogenic vulnerabilities to be generated—attempts to address Covid-19-related vulnerabilities can create new vulnerabilities or exacerbate existing ones.[44] There is, therefore, a more urgent need to identify and respond to any emergent vulnerabilities that relate to use of Big Data in this research context, as we do in the following section.

BALANCING THE KEY ETHICAL VALUES

In elucidating the key ethical implications for UK-REACH and similar research projects, we draw on the substantive ethical values identified by Xafis and colleagues' framework, as highlighted above.[45] Although all of the ethical values identified by this framework are of relevance for Big Data research, for the purposes of this article and with a specific view to public health research, we highlight three core values, namely, privacy, justice, and public benefit. We explore how these values interact, how they ought to be conceptualized for public health Big Data usage. This helps set the stage for a Big Data ethics-by-design approach to incorporate core ethical principles at the earliest stages of Big Data-driven public health research projects.

Privacy as an ethical value. Above, we have discussed privacy as both a legal right and a human right, but it is also important to recognize that privacy is of ethical significance. Data subjects have an interest in the assurance of their privacy, including use of their data in line with their morally reasonable expectations.

Just as we note that Big Data-driven public health research ought to incorporate principles that are more attuned to community-focused public health ethics rather than individualistically oriented clinical or biomedical ethics, so too do we stress that privacy as an ethical value ought to encompass concerns that transcend the paradigm of an individual's having information about them held in a state of noninterference. In our view, privacy ought to also encompass group concerns about information that can be considered morally significant for a given community and that can give rise to questions about how that information ought to be collected, used, and disclosed. In other words, for Big Data-driven public health research projects, we ascribe to a broader notion of privacy that considers *both* the individual (qua research participant or data subject) and the community or communities with which they identify and in which they are situated.

The key privacy concerns arising for projects such as UK-REACH surround the use of a broad range of data about individuals, in this instance, about their health, employment, and private lives. Taken together,

such information use can lead to important findings about segments of the U.K. population, and specifically segments of the HCW population, but that information use ought to align with reasonable expectations of privacy, on both individual and community levels.

Ensuring privacy interests are respected may at first glance appear straightforward because, as part of standard practice, data directly identifying individuals within the dataset are removed prior to its use by researchers. However, given our above-mentioned legal (and practical) distinction between anonymization, pseudonymization, and deidentification, there remains a need to consider whether the linked, anonymized dataset can truly be considered anonymous, especially where a large quantity of information about a given individual is being retained for analysis. Where this is the case, there is a potential for the data to remain identifiable because it becomes increasingly unlikely that a single combination of values within the dataset will apply to more than one individual.[46] For example, Ohm details how the combination of zip code, sex, and date of birth can be used to identify the majority of the population in the U.S.—a much larger population, of course, than in the U.K.[47] This demonstrates the surprising ease with which ostensibly "anonymized" data may in fact be reidentified. Where more detailed information is contained, the chance that a given combination is unique to one person becomes greater, and the possibility of being able to link those data to that individual is higher.

Moreover, the risk of reidentification is amplified through the linkage of previously unconnected datasets. Even if linkage of data increases utility in research, it is important to retain an awareness that the linked data will be a far richer account of a given individual,[48] revealing more about who they are and their lives than each dataset in isolation could. Linkage also further decreases the likelihood that more than one individual has all the same characteristics. Using UK-REACH to illustrate this, even if more than one person could work in the same hospital with the same job role and ethnicity (their deidentified or anonymized employment data could not be distinguished from someone else's), when we combine employment data with health outcomes, it is highly unlikely there will be more than one person with all those features *and* all the same underlying health conditions, Covid-19 status, and clinical outcome. This demonstrates that removal of direct identifiers is not in and of itself sufficient to guarantee privacy for many research projects using Big Data.[49] Therefore, we need to consider what other measures are required to protect privacy interests, not only from an ethical perspective but also for legal compliance—if data are not truly anonymous and the GDPR Recital 26 standard is not met, the GDPR will continue to be applicable even after the data are anonymized.

In highlighting the risks of reidentification for such research projects, we are not aiming to suggest that anonymization is devoid of utility. Without question, effective anonymization makes it more difficult (if not effectively impossible) to identify data subjects and thus can provide some protection for privacy interests. Rather, we propose three key modifications to the predominant understanding of anonymization in Big Data research such that it better addresses the privacy-as-ethical-value concerns we have highlighted in this article.

First, we share Mourby's perspective that anonymization should be viewed as part of a framework for ethical information governance.[50] Further measures for ensuring privacy should also be implemented, including seeing to it that the data are used (and shared) only for appropriate purposes and that their security is guaranteed (which is key in ensuring public acceptability).[51] This means controlling how data are accessed and who accesses them and allowing access only to data that is necessary for research purposes.[52]

From UK-REACH, we view the use of a trusted research environment (such as the SAIL Databank) as a vital measure that gives significant protection to privacy. Environments such as SAIL will ensure that the data enjoys heightened digital and physical security and that the researchers are given limited access to the data and only on completion of appropriate training. This ensures that the data remain confidential and out of the public domain, limits those who can access data (and requires them to be trained), and follows established and accepted pathways for data to be shared and managed throughout their lifecycle.

The approach taken to the process of deidentification and anonymization also plays an important role in ensuring privacy. Where identifiers are encrypted (given heightened security) and deleted once they are no longer required (minimizing the data to what is

necessary), this will also protect the identity of an individual. But, even with these measures, those in control of the data in question must always remain vigilant to the prospect that data might at some point in the future become identifiable.

Following from the last point, we encourage a reframing of the anonymization process to make it more dynamic than it typically is in Big Data research projects. Techniques to anonymize data should not use predetermined measures to deidentify, as these may fail to properly reflect the context of the data, including shifting contexts over time.[53] Instead, what is identifiable should be considered by reference not just to what is identifiable in most contexts but also to what is identifiable in conjunction with data that are readily available or may be available to anyone seeking to reidentify data.[54] This becomes particularly important when the findings of research are reported and disseminated. It is vital to ensure that the data presented in the findings cannot be reidentified once they are publicly available (and therefore subject to less control).

Finally, anonymization should not be viewed as a one-off event (so-called release-and-forget anonymization).[55] A watching brief is required to ensure the data remain anonymized throughout their lifecycle in the research endeavor and that the risk of data triangulation is mitigated. Part of the aim of data linkage is to provide resources to future studies, and it is especially important if data are being reused that the privacy of information is kept under close review. For UK-REACH, there is a temptation to conclude that because the data are held in the secure SAIL Databank, this obligation is met and anonymization is not a significant concern. However, to fully adopt a dynamic approach to anonymization requires that UK-REACH go beyond this and actively seek to test whether the data being made available to researchers are anonymous, undertaking checks throughout the project to do so. This would allow the project to better ensure that it is protecting the privacy of its data subjects.

This more nuanced and dynamic approach also serves our broader notion of privacy and allows for consideration of communitarian privacy interests, a notion spelled out in more detail by Floridi, among others.[56] By using what we refer to as "dynamic anonymization," research projects are able to consider if certain pieces of information are of specific relevance to certain populations and therefore might still be considered identifiable by reference to that group. Here, we adopt Loi and Christen's two conceptions of group privacy to determine identifiability, which can be thought of as two poles in a continuum: (1) groups consisting of natural persons with an interaction history and/or collective goals in the sense of displaying some meaningful form of agency, as a group, for example, through intentional coordination, or at least awareness of themselves as a group with which they identify, and (2) groups consisting of natural persons with one or more features in common who do not have an awareness of or stake in setting aside the trivial case of shared goals, which are pursued without a common plan or for the common good (e.g., smokers share the goal to smoke).[57] Loi and Christen argue that the second group, which gives rise to inferential privacy concerns, deals with the inferences that can be made about a group of people defined by a feature, or combination thereof, shared by all individuals in the group. Big Data analytics is especially threatening to the inferential privacy of individuals that are characterized by features common to open-ended groups. Floridi cautions that "[w]e need to be more inclusive [to group privacy interests] because we are underestimating the risks involved in opening anonymized personal data to public use, in cases in which *groups* of people may still be easily identified and targeted."[58] In our view, adopting a dynamic approach to anonymization would, at a minimum, better position researchers to mitigate identifiability (and stigmatization) risks to individuals and inferences that can be made about the groups within which they are situated, including inferences about their health status or predisposition to disease.

It is equally important that projects consider that anonymization is not undertaken to such a degree that data become inappropriately homogenized and the interests of certain communities are obscured. Where dynamic anonymization throws up such communitarian privacy concerns, engagement with those relevant communities is crucial to ensure that researchers avoid any assumptions about which characteristics may be important and that the dynamic anonymization process is undertaken in a way that supports their privacy interests in practice. Such an approach will allow projects to strike the appropriate balance between maintaining the utility

of the dataset and ensuring that the privacy interests of participants and communities can be respected.

Thus, we view dynamic anonymization (as part of a framework for good information governance) as a key aspect of the Big Data ethics-by-design approach for any Big Data project. In practice, this framework requires each project to consider what privacy requires for the project, at its outset, throughout its duration, and in the dissemination and use of its findings, including active interrogation of the data made available to researchers throughout the research project's lifecycle. This approach allows anonymization to retain its utility as a method (but not *the* method) of protecting privacy interests.

Justice as an ethical value. Justice should be at the heart of responding to vulnerabilities, for both Big Data and public health research. It requires, at a minimum and in the negative sense, that data are not used in a way that exacerbates discrimination against and power asymmetries between different groups of health professionals, and different genders and ethnicities in U.K. society (mirroring the need for public health measures to not be undertaken in a manner that exacerbates health inequalities). In the positive sense, data should be used to provide greater distributive justice through making hitherto neglected or invisible communities more visible. More generally, due regard should be given to fair access, participation, and representation in the project so that those most vulnerable to discrimination and invisibility have an opportunity to contribute meaningfully to the project's design, delivery, and dissemination.

For a project like UK-REACH, it is particularly crucial to address any issues of inequity arising from research. The focus of this research is ethnic minority groups who already experience structural inequality, marginalization, and discrimination; it is therefore necessary to mitigate any potential pathogenic vulnerabilities that might arise from the research and dissemination activities (including policy translation). From a data perspective, ethnic minority groups have traditionally (especially in the health context) had unequal access to Big Data and research or been excluded from its benefits.[59] Taking active steps to ensure that this group both participates in and benefits from the research is thus necessary. Such steps include targeted recruitment of ethnic minority HCWs and the involvement of minority ethnic groups as stakeholders from research design through to the translation of findings into policy.

A second key aspect of justice is to give due consideration to the possibility that research outputs could result in unintended pathogenic vulnerabilities and group harms. This might include heightened social stigma arising from the dissemination of findings on ethnic minority HCWs, or policies aimed to protect vulnerable groups that result in unfair differential treatment. For example, should particular health care roles, such as working in intensive care units or intensive therapy units, be considered "too risky" for ethnic minorities because of Covid-19, this could negatively affect choices available to ethnic minority individuals in terms of career pathways, including limiting trainee doctors' opportunities for progression. Instead, if evidence does indeed suggest heightened concern, providing additional support and protections (in ways that support and foster agency)[60] should be considered as a means to prevent harm.

Second, given existent unequal access and misuse of data involving ethnic minority groups, there may be heightened concern about potential misuse of data and mistrust in bodies handling them. This anxiety is likely to increase in a pandemic that has seen such a disproportionate impact on ethnic minority groups. Key to addressing this and meeting the requirements of justice is for such projects to seek to broach the "Big Data divide" and give opportunities for these groups (and the subjects of research generally) to participate in and influence discussions around how their data are used in research. Careful thought should be given as to how to engage various groups to ensure appropriate representation, as well as how to capture and reflect the values of groups and feed those values into decision-making processes in research. It is equally important, given the diversity among people, that groups are not homogenized or findings generalized in a manner that fails to recognize relevant and intersectional experiences.

What justice requires for different public health projects will depend very much on their particular context, and as UK-REACH focuses on ethnicity and HCWs, the analysis above naturally centers on those concerns. A Big Data ethics-by-design approach would require any project, at a minimum, to identify any groups relevant to the research that experience existing

structural inequality or power imbalance and to consider how these groups may be included and benefited. This includes consideration as to what group harms could arise and how these should be mitigated. This understanding of justice requires research projects to actively discuss justice issues that come up during the project and consider the translation of their research findings into policy. UK-REACH could do this on a more active basis, for instance, by considering throughout its data collection and analysis stages how the data could be misused and in what ways use of those data and downstream results could disproportionately affect certain groups. Although research projects have a limited lifespan (and thus may be unable to fully control what is done with their findings and downstream uses), researchers arguably have a moral duty to question the instrumental goals and the individual timelines of specific projects and to design into their research longer-term forms of public participation in science. This would entail an interrogation of what the goals of the project are in the context of justice and how this ought to be folded into the design of the project and beyond its duration. If the purpose of undertaking the research is to inform policy, then the timeline for the project should include the nurturing of relationships with policy-makers at an early stage, as well as a period to engage with them to discuss how the research is translated into policy and prevent any measures that may lead to discriminatory outputs or group harms.

Across the duration of the project and translation of findings, we consider that engagement with stakeholders and those with morally relevant interests will allow projects to mitigate the risk of inappropriate impacts arising from research. In UK-REACH, this is achieved through WP5 interactions with the stakeholder engagement groups, which help ensure that groups with morally relevant interests are given a clear opportunity to inform how the research is conducted, raise any concerns, and shape its translation into policy. This approach could be strengthened further, however, by actively raising any concerns about misuse of the data that researchers have identified with the stakeholder groups so that these groups can also provide input for the mitigation of any potential harms.

We also stress that the nature of stakeholder engagement must be meaningful.[61] Mere participation in discussions, or the existence of a de facto stakeholder group, would not be sufficient to satisfy the requirements of justice if the views of stakeholder groups had no influence or impact on how research is conducted or how findings influence policy. It is vital that the researchers act upon the concerns raised by stakeholders, where appropriate.

Balancing privacy and justice with the public interest of research. As we have highlighted above, UK-REACH is a project at the intersection of public health and Big Data research. Because this is a public health project, it is necessary to consider and interrogate what benefits are likely to arise from the study and, recognizing the contextualized nature of benefit in this area of research, what this means for various publics. As we have noted, this is not required just from an ethical perspective but also to support the legal basis for data processing by UK-REACH. Important UK-REACH benefits for publics might include immediate answers relating to the clinical outcomes of Covid-19 in ethnic minority HCWs, a better evidence-based understanding of Covid-19 that can inform responses to current and future pandemic waves, and a framework within which researchers and policy-makers are able to investigate longer-term clinical sequelae (on physical health and mental health). However, it is important that UK-REACH continues to be reflexive about (1) the purported benefits, interrogating how the research is in the public interest, and about (2) which of these potential benefits may be realized and (3) how they may be realized in an ethical manner. This should be done on a more proactive basis to fully meet the requirements of Big Data ethics by design.

Beyond the work of articulating and interrogating the benefits (to ensure they are indeed realized), consideration should be given to how to balance the purported benefits with the interests of privacy and justice we have highlighted. Each value has great importance in its own right. Privacy concerns arise whenever individual data are used (and are more pronounced, arguably, when they involve health data), but in the public health context, the public benefit of their use may well outweigh individual interests (unlike in clinical and other research contexts, which may give priority to individual rights). Equally, concern around Covid-19's disproportional impact on

ethnic minorities means that justice needs to be central to this balancing act.

Therefore, we consider that finding this balance should be rooted in proportionality and harm minimization,[62] through a bespoke approach for any given project. This requires projects to articulate and continue to reflect upon the purported benefits of the research and to ensure that use of data, and any privacy concerns that come up, remain appropriate to achieve this benefit. Particular attention should be paid to issues of justice, with consideration of any potential group harms or inequity arising from the gathering, sharing, and use of participant data. Such an approach will also support legal compliance of the project, should the legal basis for data processing be article 6(1)(e) and/or article 9(1)(i) of the GDPR.

In considering what constitutes an appropriate and proportionate balance, Xafis and colleagues' procedural values provide much insight.[63] Engagement with the principal stakeholders in a research project is particularly crucial. This ensures those with morally relevant interests can influence how the research develops, which in turn helps ensure that the benefit for relevant publics is achieved and assists with the identification and appropriate minimization of potential group harms. Also important is to demonstrate the trustworthiness of the project through openness to public scrutiny and transparency of processes throughout the project's lifetime.

Using this approach, projects such as UK-REACH can ensure the ethical acceptability of the project, even in the uncertainty caused by the Covid-19 pandemic. Taking an engaged, transparent, and reflexive approach will also allow the project to demonstrate what ongoing measures are being taken to protect participants' data and interests, as well as enable response to any pandemic-specific concerns that are raised (indeed, this is arguably more important in a pandemic where rapid research is increasingly common).[64] Such an approach will also allow any new challenges or hitherto unknown unknowns to be rapidly and properly considered to see to it that harms are identified and minimized where required.

Potential measures we highlight to meet this ethical approach are as follows:

- **Stakeholder involvement throughout research.** Devising a public involvement strategy, such as that reflected in WP5 of UK-REACH, is vital. This will ensure there is meaningful and constructively critical engagement with stakeholders—and research investigators—throughout the research project, all the way through policy translation. It will ensure that decisions made are transparent and contribute to making the research trustworthy. By engaging with multiple publics, the project will also be better able to interrogate and assess whether it is meeting the proposed public benefit.
- **Use of a trusted research environment.** A trusted research environment such as the SAIL Databank gives heightened protection to individual privacy without minimizing utility and benefit. The additional protections that come with such an environment ensure the use of data is limited to what is necessary. Such an environment will also serve to increase public acceptability and trust by assuaging concerns or doubts about misuse of data. Continuing to reflect on what information is held in the relevant trusted research environment and whether this is necessary in meeting the expected benefits of the research remains important.
- **Provision of public-facing information.** Research findings must be accessible and open to scrutiny if they are to be seen as beneficial and meet the interests of justice. Projects such as UK-REACH ought to provide information about research activities on an ongoing basis. This information should be publicly accessible and easy to find and should include the aims and methods of the research project, as well as the findings. Dissemination through other forms of media (podcasts, short videos, blogs, and so on), including social media, ought to be considered as they can further reach publics that might not otherwise be engaged. Where the aim is to inform policy, it is also important to nurture relationships with policy-makers and to make the findings available to them in a timely manner. This allows policy decisions to be made in near real time and with as much evidence as possible at hand.

CONCLUSION

In this article, we undertook a two-part examination of the UK-REACH project, first identifying the core legal and ethical issues that this project raises (focusing in particular on ethical issues) and then considering ways in which those issues might be addressed to advance positive aspects while mitigating or eliminating

any negative aspects. Adopting Xafis and colleagues' "Ethics Framework for Big Data in Health and Research," we looked at how privacy and justice interact and how we should conceptualize this for public health Big Data usage. Ultimately, we argued that taking an engaged, transparent, and reflexive approach will yield the most benefit for projects of this type. We also argued that through active involvement of stakeholders, use of a trusted research environment, and provision of public-facing information (that remains up to date), Big Data research projects, especially those operating in the midst of a public health emergency (if not a pandemic), will more likely be viewed as ethically robust and as yielding research insights that are deserving of policy uptake to improve the health and well-being of populations across the globe. As this article has demonstrated, the current era of large-scale, data-driven health research projects (which the Covid-19 pandemic has shown significantly help drive scientific breakthroughs that translate into medical innovations and effective public health interventions and which are now conducted in many countries around the globe) raise a number of ethico-legal challenges. Adopting a Big Data ethics-by-design approach when constructing such projects helps meet those challenges. This approach enables robust identification of and efficient and effective responses to ethical concerns that are raised in these projects and also enables any new challenges or unknown unknowns to be rapidly and properly considered to ensure harms are identified and addressed where required. ♦

SUPPORTING INFORMATION

The appendix is available in the "Supporting Information" section for the online version of this article and via *Ethics & Human Research*'s "Supporting Information" page: https://www.thehastingscenter.org/supporting-information-ehr/.

Ruby Reed-Berendt, LLB, LLM, *is a research associate with the UK-REACH project (work package 3) and a PhD candidate at the School of Law at the University of Edinburgh;* **Edward S. Dove, PhD,** *is a lecturer in health law and regulation at the School of Law at the University of Edinburgh; and* **Manish Pareek, MBChB (Hons), PhD, MRCP,** *is an associate clinical professor of infectious diseases in the Department of Respiratory Sciences at the University of Leicester and an honorary consultant in infectious diseases in the Department of Infection and HIV Medicine at the University Hospitals of Leicester NHS Trust.*

ACKNOWLEDGMENTS

The authors would like to thank Catherine Montgomery and Graeme Laurie for their comments on a previous draft.

DISCLOSURE

This study is supported by a grant (MR/V027549/1) to the University of Leicester from the Medical Research Council-U.K. Research and Innovation and the National Institute for Health Research (NIHR) rapid response panel to tackle Covid-19 and by core funding provided by NIHR Leicester Biomedical Research Centre, a partnership between the University of Leicester and University Hospitals of Leicester NHS Trust. This work is carried out with the support of BREATHE—Health Data Research Hub for Respiratory Health (with grant MC_PC_19004) funded through the U.K. Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research U.K. MP is funded by an NIHR Development and Skills Enhancement Award.

REFERENCES

1. See e.g. Conroy, M., et al., "The Advantages of UK Biobank's Open-Access Strategy for Health Research," *Journal of Internal Medicine* 286, no. 4 (2019): 389-97; Beesley, L., et al., "The Emerging Landscape of Health Research Based on Biobanks Linked to Electronic Health Records: Existing Resources, Statistical Challenges, and Potential Opportunities," *Statistics in Medicine* 39, no. 6 (2020): 773-800; Shilo, S., H. Rossman, and E. Sega, "Axes of a Revolution: Challenges and Promises of Big Data in Healthcare," *Nature Medicine* 26, no. 1 (2020): 29-38.

2. See e.g. Pareek, M., et al., "Ethnicity and COVID-19: An Urgent Public Health Research Priority," *Lancet* 395 (2020): 1421-22; Lusignan, S., et al., "Risk Factors for SARS-CoV-2 among Patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre Primary Care Network: A Cross-Sectional Study," *Lancet Infectious Diseases* 20 (2020): 1034-42; Pan, D., et al., "The Impact of Ethnicity on Clinical Outcomes in COVID-19: A Systematic Review," *EClinicalMedicine* 23 (2020): doi:10.1016/j.eclinm.2020.100404.

3. Patel, P., et al., "Ethnicity and Covid-19," *BMJ* 369 (2020): doi:10.1136/bmj.m2282.

4. "The United Kingdom Research Study into Ethnicity and COVID-19 Outcomes in Healthcare Workers," UK-REACH, November 21, 2021, https://uk-reach.org/main/.

5. See Gogoi, M., et al., "Ethnicity and COVID-19 Outcomes among Healthcare Workers in the United Kingdom: UK-

REACH Ethico-Legal Research, Qualitative Research on Healthcare Workers' Experiences, and Stakeholder Engagement Protocol," *medRxiv* (2021): doi:10.1101/2021.03.03.21252737.

6. Ibid.

7. "SAIL Databank," November 21, 2021, https://saildatabank.com/.

8. Fernandez Lynch, H., D. Lundin, and E. A. Meagher, "Ethical Inclusion of Health Care Workers in Covid-19 Research," *Ethics & Human Research* 43, no. 2 (2021): 19-27.

9. See generally, Cohen, I. G., et al., eds., *Big Data, Health Law, and Bioethics* (New York: Cambridge University Press, 2018).

10. See, e.g., Iphofen, R., and M. Kritikos, "Regulating Artificial Intelligence and Robotics: Ethics by Design in a Digital Society," *Contemporary Social Science* 16, no. 2 (2021): 170-84; Urquhart, L. D., and P. J. Craigon, "The Moral-IT Deck: A Tool for Ethics by Design," *Journal of Responsible Innovation* 8, no. 1 (2021): 94-126.

11. 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), OJ 2016 L 119/1, at art. 4.

12. Campbell v. MGN Limited (2004) UKHL 22. See also AG v. Guardian Newspapers (No. 2) (1990) 1 AC 109 at 281.

13. X v. Y (1988) 2 All ER 648; Hunter v. Mann (1974) QB 767; Ashworth Hospital Authority v. MGN (2002) UKHL 29.

14. See General Medical Council, *Confidentiality: Good Practice in Handling Patient Information* (2017), https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/confidentiality, and Nursing and Midwifery Council, *The Code: Professional Standards of Practice and Behaviour for Nurses, Midwives and Nursing Associates* (2018), https://www.nmc.org.uk/standards/code/.

15. Vidal-Hall v. Google Inc (2015) EWCA Civ 311.

16. Coco v. AN Clark (Engineers) Ltd (1968) FSR 415.

17. European Court of Human Rights, Council of Europe, European Convention on Human Rights, 2021, https://www.echr.coe.int/documents/convention_eng.pdf.

18. Vidal-Hall v. Google Inc.

19. According to article 4(5) of the GDPR, "pseudonymisation" means "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."

20. Regulation (EU) 2016/679 of the European Parliament, recital 26.

21. Ibid.

22. S and Marper v. United Kingdom (30562/04) (2008) WLUK 117.

23. LH v. Latvia (2015) 6 EHRR 17.

24. Article 8(2) of the European Convention on Human Rights states, "There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others."

25. Xafis, V., et al., "An Ethics Framework for Big Data in Health and Research," *Asian Bioethics Review* 11 (2019): 227-54.

26. Nuffield Council on Bioethics, *The Collection, Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues* (Nuffield Council on Bioethics, 2015), https://www.nuffieldbioethics.org/publications/biological-and-health-data, pp. 23-33.

27. Ibid., pp. 33-42.

28. National Health Services, Health Research Authority, "COVID-19 Research," November 21, 2021, https://www.hra.nhs.uk/covid-19-research/.

29. Dawson, A., and B. Jennings, "The Place of Solidarity in Public Health Ethics," *Public Health Reviews* 34 (2012): 4.

30. Coggon, J., and A. M. Viens, "Public Health Ethics in Practice: A Background Paper on Public Health Ethics for the UK Public Health Skills and Knowledge Framework," Public Health England, April 21, 2017, https://www.gov.uk/government/publications/public-health-ethics-in-practice.

31. Xafis et al., "An Ethics Framework for Big Data in Health and Research."

32. Mackenzie, C., W. Rogers, and S. Dodds, eds., *Vulnerability: New Essays in Ethics and Feminist Philosophy* (Oxford: Oxford University Press, 2013).

33. Andrejevic, M., "The Big Data Divide," *International Journal of Communication* 8 (2014): 1673-89.

34. Xafis et al., "An Ethics Framework for Big Data in Health and Research."

35. Loring, B., and A. Robinson, "Obesity and Inequities: Guidance for Addressing Inequities in Overweight and Obesity," World Health Organization Regional Office for Europe, 2014, https://www.euro.who.int/__data/assets/pdf_file/0003/247638/obesity-090514.pdf.

36. Xafis et al., "An Ethics Framework for Big Data in Health and Research."

37. See generally, Slovic, P., *The Perception of Risk* (New York: Routledge, 2000).

38. Denham, E., "Data Protection Considerations and the NHS COVID-19 App," *ICO.* (blog), September 18, 2020,

https://ico.org.uk/global/data-protection-and-coronavirus-information-hub/blog-data-protection-considerations-and-the-nhs-covid-19-app/.

39. Montgomery, C., and R. Pool, "From 'Trial Community' to 'Experimental Publics': How Clinical Research Shapes Public Participation," *Critical Public Health* 27, no. 1 (2017): 50-92.

40. Government Equalities Office, Race Disparity Unit, and Badenoch, K., "Quarterly Report on Progress to Address COVID-19 Health Inequalities," Government of the United Kingdom, October 22, 2020, https://www.gov.uk/government/publications/quarterly-report-on-progress-to-address-covid-19-health-inequalities.

41. Ganguli-Mitra, A., "The Need to Unpack Vulnerability in a Pandemic," *BMJ Opinion* (blog), July 7, 2020, https://blogs.bmj.com/bmj/2020/07/03/agomoni-ganguli-mitra-the-need-to-unpack-vulnerability-in-a-pandemic/.

42. Patel et al., "Ethnicity and Covid-19."

43. Mahase, E., "Covid-19: Many ICU Staff in England Report Symptoms of PTSD, Severe Depression, or Anxiety, Study Reports," *BMJ* 372 (2021): n108.

44. Mackenzie, Rogers, and Dodds, *Vulnerability*.

45. Xafis et al., "An Ethics Framework for Big Data in Health and Research."

46. Ohm, P., "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," *UCLA Law Review* 57 (2010): 1701-69.

47. Ibid.

48. Ibid.

49. Nuffield Council on Bioethics, *The Collection, Linking and Use of Data in Biomedical Research and Health Care*.

50. Mourby, M., "Anonymity in EU Health Law: Not an Alternative to Information Governance," *Medical Law Review* 28, no. 3 (2020): 478-501.

51. Davidson, S., et al., *Public Acceptability of Cross-Sectoral Data Linkage: Deliberative Research Findings* (Scottish Government Social Research, 2012).

52. These measures are reflective of the principles underpinning the GDPR (see article 5). However, we argue that there is an ethical requirement to protect the data in a manner that ensures privacy, beyond the requirements of the GDPR.

53. Nuffield Council on Bioethics, *The Collection, Linking and Use of Data in Biomedical Research and Health Care*.

54. Ibid.

55. Ohm, "Broken Promises of Privacy."

56. Floridi, L., "Group Privacy: A Defence and an Interpretation," in *Group Privacy: New Challenges of Data Technologies*, ed. L. Taylor, L. Floridi, and B. Van der Sloot (Cham, Switzerland: Springer, 2016), 83-100.

57. Loi, M., and M. Christen, "Two Concepts of Group Privacy," *Philosophy & Technology* 33 (2020): 207-24.

58. Floridi, "Group Privacy," 98

59. Boyd, K., "Ethnicity and the Ethics of Data Linkage," *BMC Public Health* 7 (2007): 318.

60. McKenzie, "The Importance of Relational Autonomy and Capabilities for an Ethics of Vulnerability," in *Vulnerability*, 33-59.

61. Reynolds, L., and S. Sariola, "The Ethics and Politics of Community Engagement in Global Health Research," *Critical Public Health* 28, no. 3 (2018): 257-68.

62. Xafis et al., "An Ethics Framework for Big Data in Health and Research."

63. Ibid.

64. Ratneswaren, A., "The I in COVID: The Importance of Community and Patient Involvement in COVID-19 Research," *Clinical Medicine Journal* 20, no. 4 (2020): e120-22.

ERRATUM

In the acknowledgment section for the article "HIV Cure Research: Risks Patients Expressed Willingness to Accept," by Allison Kratka, Peter A. Ubel, Karen Scherr, Benjamin Murray, Nir Eyal, Christine Kirby, Madelaine N. Katz, Lisa Holtzman, Kathryn Pollak, Kenneth Freedburg, and Jennifer Blumenthal-Barby (*Ethics & Human Research* 41, no. 6 [2019]: 23-34, doi:10.1002/eahr.500035), some funding information was accidentally omitted. The authors also acknowledge funding from the National Institute of Allergy and Infectious Diseases through grant R01 AI114617 (HIV Cure Studies: Risk, Risk Perception, and Ethics, with Eyal as the principal investigator). DOI: 10.1002/eahr.500113