

Asserting the public interest in health data: On the ethics of data governance for biobanks and insurers

Big Data & Society
October–December: 1–12
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517241290215
journals.sagepub.com/home/bds



Kathryne Metcalf¹  and Jathan Sadowski² 

Abstract

Recent reporting has revealed that the UK Biobank (UKB)—a large, publicly-funded research database containing highly-sensitive health records of over half a million participants—has shared its data with private insurance companies seeking to develop actuarial AI systems for analyzing risk and predicting health. While news reports have characterized this as a significant breach of public trust, the UKB contends that insurance research is “in the public interest,” and that all research participants are adequately protected from the possibility of insurance discrimination via data de-identification. Here, we contest both of these claims. Insurers use population data to identify novel categories of risk, which become fodder in the production of black-boxed actuarial algorithms. The deployment of these algorithms, as we argue, has the potential to increase inequality in health and decrease access to insurance. Importantly, these types of harms are not limited just to UKB participants: instead, they are likely to proliferate unevenly across various populations within global insurance markets via practices of profiling and sorting based on the synthesis of multiple data sources, alongside advances in data analysis capabilities, over space/time. This necessitates a significantly expanded understanding of the publics who must be involved in biobank governance and data-sharing decisions involving insurers.

Keywords

Risk, trust, de-identification, machine learning, actuarial science, algorithmic discrimination

Introduction

In November 2023, investigative reporting by *The Observer* revealed that the UK Biobank (UKB)—a publicly-funded organization which aggregates health and genetic data from over half a million UK citizens—had shared sensitive health data with a variety of insurance firms (Das, 2023). The story included outraged commentary from health researchers and data ethicists, who described the incident as a “disturbing breach of trust” meriting serious reconsideration of the organization’s data access policies and public communications.

In return, a heated response by UKB Principal Investigator and Chief Executive Rory Collins—released the same day as the *Observer* story—described the reporting as a “highly misleading... false narrative” (Collins, 2023), though he stopped short of identifying significant factual error in the piece. Instead, Collins’ took issue with how the data-sharing practices were characterized as socially problematic or institutionally deceptive. Rather than apologizing or denying the allegations, Collins sought to justify the UKB’s practices as legitimate by emphasizing three points: (a) that the shared data are

de-identified, (b) that participants gave consent for commercial research, and (c) that actuarial projects conducted by private insurers were aligned with the UKB mandate to produce “health-related research which is in the public interest.” He concluded by describing the reporting as “irresponsible,” “extremely disappointing,” and “highly regrettable,” alleging that *The Observer* itself was to blame for any lapse of public trust in the UKB that might result from its data-sharing disclosures.

Despite the somewhat theatrical tone of this exchange, the incident has not (yet) blown over into larger controversy. However, as scholars of biobanking and insurance technologies, this incident appears to us as both a

¹Department of Communication, University of California San Diego, La Jolla, CA, USA

²Emerging Technologies Research Lab, Faculty of Information Technology, Monash University, Melbourne, VIC, Australia

Corresponding author:

Kathryne Metcalf, University of California San Diego, La Jolla, CA, USA.
Email: kymetcal@ucsd.edu

warning of things to come and indicative of a broader set of concerns related to the politics and ethics of health data governance. While the research projects for which actuarial firms requested UKB data vary in their specifics, all of them proposed to develop tools using detailed health data from the biobank to assess disease risk and predict morbidity outcomes. Although this data does have the potential to contribute to transformative improvements in healthcare, we argue there is little reason to believe that such benefits will be realized—much less in any substantial or pro-social way—via private sector actuarial research. Collins’ claim relies on a specious assumption that the public’s interest will be served by the spillover or trickle-down effects of corporate actuarial research; however, the core logic of the insurance industry focuses on developing techniques to identify, manage, and eliminate any (potential) risk that is not outweighed by the possibility to realize value (Sadowski, 2023). In the provision of health insurance, then, the identification of new health risks simultaneously produces new financial risks an insurer must seek to minimize—and they do so in ways which can have significantly negative impacts for consumers.

Tracing the implications of this point, we argue that the use of biobank data for insurance research raises crucial questions for bioethics, data ethics, and the governance of publicly-funded data resources. Principally, we seek to contest the implications of two of Collins’ claims: that insurance research is “in the public interest,” and that de-identification meaningfully negates individuals’ stake in how their data are used. Examining the UKB as our primary case study, this paper makes two interlocking contributions. First, it offers a critical analysis of how institutions entrusted as stewards of sensitive data use (and abuse) ethico-political concepts like “the public interest”—which they imbue with ambiguous and idiosyncratic meaning—to justify practices that might otherwise be seen as problematic. We offer alternative understandings of the public interest more in line with popular opinion based on recent survey data, and draw out what they would mean when taken seriously as governance principles. Second, this paper offers a normative argument about the failures of current health data governance practices to take seriously the relationality of data, and considers the responsibilities such institutions have if they are to be considered trusted stewards.

The paper proceeds as follows. In the second section, we provide general background on biobanks and the valuation of biobank data along with more specific details about our motivating case study. In the third section, we show that claims about the “public interest” are used to simultaneously justify and conceal the moral economy of actuarial governance, and we make the case that actuarial research intended for commercial deployment by insurers is *not* in the public interest, nor does the public see it as in their interest. As we argue, insurer rhetoric about medical data

conflates the potential for research-driven beneficial health outcomes with the research-driven prioritization of profitable risk selection by insurers. In the fourth section, we problematize the equivalence of de-identification with meaningful protection for UKB research participants and nonparticipants alike. We outline how the logic of insurance necessarily relies on market discrimination against “risky” groups—groups which the insurance technologies developed with UKB data aim to identify—who can then be legitimately profiled through the use of data accumulated by other means. We also examine other cases where countervailing group interests have demanded alternative health data governance frameworks, suggestive of how UKB data-sharing policies could be constructed otherwise. In concluding, we further consider broader normative points about the failures and opportunities for data governance that secures the public interest against private desires.

From biobanking to (bio)capital

The UKB is a prominent example of an increasingly expansive set of institutions called biobanks: organizations which aggregate large collections of biomaterials (like saliva or blood) and related data intended for research use. While estimates vary widely, one recent study suggests that there are as many as 30 biobanks per million residents across countries with high research productivity (O’Donoghue et al., 2022), and new collections are founded every year. These organizations run the gamut from small hospital-maintained pathology libraries and private nonprofits to massive population registers integrated within national health systems. Although the majority of biobanks receive minimal public attention, their scientific impact is profound: biobanking has made new data-intensive research methods (like Genome-Wide Association Studies) not simply *possible* but *common* across the life sciences, and has accelerated the development of innumerable diagnostics and therapeutics. Even as we write this essay, an article describing newly available data from a major biobank—the All of Us Project, an American cognate to the UKB—is among the most-read new publications in *Nature* (Bick et al., 2024).

Across the large and diverse landscape of contemporary biobanking, the UKB is notable for its sheer size—both in terms of its number of research subjects and the amount of data collected about each. The resource has enrolled half a million UK participants who provide multiple biomaterial and genetic samples, complete comprehensive surveys about their lifestyle and behavior, link their National Health Service medical records, and submit to various forms of testing, imaging, and other measurements (Littlejohns et al., 2019; Bycroft et al., 2018).^[1] These data are significantly intimate: they provide a detailed record of each participant’s health and life-course, far exceeding the type of information you might share in a routine visit to a doctor or one-off research study. The

UKB has developed a specialized Ethics and Governance Framework (2007) and retains an independent Ethics and Governance Council in recognition of this sensitivity. Its approach to governance is considered a model for many biobanking organizations worldwide, such that ethically-sensitive data-sharing decisions made by the UKB can be understood to set an important precedent.

Importantly, biobanks have long been tethered to contestations about the value of their materials: the “bank” metaphor in biobanking is not incidental (Metcalf, 2022). Indeed, much of the early scholarship on biobanks in Science and Technology Studies (STS) and related fields focused on the question of value, examining how biobanks work to construct novel assemblages of private financing and governmental resources to trade in tissues and data. For instance, Mike Fortun’s (2008) pioneering account of the Icelandic biobank deCODE highlights how such organizations rely on a promissory rhetoric of value soon-to-be-realized through financial speculation in novel medical technologies. Similarly, Mitchell and Waldby (2006) highlight how the types of tissue economies in which biobanks participate produce a uniquely permeable border between symbolic discourses of the body and financial exchange in formal markets. This corpus became an important set of contributing evidence within developing theories of *biocapital* (see Helmreich, 2008 for a review of this term’s use). Bridging Foucauldian biopolitics with Marxian political economy in the postgenomic aughts, accounts of biocapital sought to disentangle the increasingly complex imbrications of speculative finance and developing biotechnologies. As Sunder Rajan (2006) critically observes, biocapital works to confuse, conflate, and ultimately co-constitute economic and ethical values. Simultaneously, it erodes the boundaries between university research and corporate R&D, collapsing imaginations of publicly-funded science and for-profit tech development.

It is interesting to observe that more contemporary STS scholarship on biobanking is somewhat less interested in these types of political-economic questions, and that biocapital itself appears to have fallen out of theoretical vogue. We can make some guesses as to why this is the case: biomaterials are less likely to travel than data *about* biomaterials, and indeed, questions about data (particularly “free,” open-access data) motivate many recent investigations (e.g., Leonelli, 2016; Strasser, 2019). Simultaneously, the speculative financial imaginations that once seemed confined to biotechnology now inform financialization processes broadly—Birch and Tyfield (2013) even make the case that work on biocapital wrongly assumed that it was innovations in the life sciences, rather than broader transformations of capital itself, that engendered these markets.

Nevertheless, we think there is something in this older work worth returning to—particularly in light of the conflation of speculative economic value (here, for private insurers) and socio-ethical value (here, “the public interest”)

at play in the UKB’s response. We need not insist on biocapital as a unique market form for this to be the case, nor biobanks as a unique set of actors: on the contrary, the use of population data to underwrite risk long predates both. However, as we will show, actuarial technologies built on AI/ML, coupled with the intensely detailed personal data accumulated by biobanks like the UKB, are poised to rapidly accelerate already-begun processes which will aggravate health disparities and worsen inequality. This demands an urgent accounting of the values—in both senses of the word—at play.

What publics? And whose interests?

While the UKB is nominally a private entity—it is legally structured as both a charity and limited company—its receipt of significant public funding, close collaboration with the National Health Service, and biosocial identification with the UK populace durably tie the organization to “the public interest,” whatever we might take that to mean. Indeed, as Collins’ letter notes, the UKB’s chief mandate is to make its data freely available “to all bona fide researchers for all types of health related research that is in the public interest” (see UKB Ethics and Governance Framework [2007] and UKB Access Policy [2022] for where this phrase appears in guidance documents). As Benjamin Capps has also observed, however, what is “in the public interest” is not a self-evident shibboleth for the UKB: its invocation opens a variety of difficult legal (2013) and bioethical (2012) considerations about appropriate data access. Here, we briefly consider the trajectory of this term as well as the role it’s played in previous debates about health data access and use. Then, we contest Collins’ framing that insurance research is consistent with the public interest.

First, though, it is clear that “public interest” is a uniquely thorny orienting concept, and it bears some reflection on why this is the case. Since the 1950s, social scientists have simultaneously figured the public interest as a north star while vigorously debating its very definition (Galston 2007)—often converging only to agree that governments should play some role in ensuring it (Dahl and Lindblom, 1954; Souraf, 1957; Schubert, 1958; Downs, 1962). However, skeptical scholarship of the next few decades questioned the stability of a singular “public” as a category (Cochran, 1974; Douglass, 1980; Mahoney, McGahan, Pitellis, 2009), as well as the ability to distinguish between real “interests” and less politically salient “desires” (Croteau and Hoynes, 2006). Even as both halves of the “public interest” seem to crumble under any scrutiny, however, those concerned with fair governance of all types continue to repeat Walter Lippmann’s famous axiom: “The public interest may be presumed to be what men would choose if they saw clearly, thought rationally, and acted disinterestedly and benevolently” (Lippmann, 1955). Similar

ideas underpin some of the most forceful theories of liberal justice, such as John Rawls' (1971) concept of “the veil of ignorance.” It remains to be seen, however, how we might all come to some agreement about what it means to see clearly or to think rationally—to say nothing of how we might conceive of (dis)interest.

Perhaps unsurprisingly, work on the public interest in health data has followed a similar trajectory—though with the added wrinkle that “health” (what it is, what are our individual responsibilities toward maintaining it, what is our interest in the health of others) is an equally unsteady category with a history of its own (Stevens, 1998). Nevertheless, the public interest has remained a prominent feature across discourses in health data governance. Writing about biobank policy in the European Union, Santa Slokenberga (2021) notes that the General Data Protection Regulation (GDPR) positions the public interest as the most important orienting principle for scientific data management. Yet, as Slokenberga also notes, “In the GDPR, public interest is mentioned 70 times, yet on none of these occasions is the concept fully explained.” The UKB no longer falls under the GDPR’s purview: post-Brexit, UK institutions are governed by the largely similar 2018 Data Protection Act. That policy, however, does not provide any further clarity. In the absence of a clear-cut definition, then, we turn to popular opinion to provide some commonsense starting points about what the public’s interest in health data might be understood to include.

Thankfully, a large body of scholarship has explored what sorts of interests and concerns the public holds about health data sharing. UK-based surveys have routinely reaffirmed that health records are broadly understood to be the most sensitive type of personal data, and have found widespread discomfort when private companies are afforded access (Hartman et al., 2020; Aitken et al., 2018; Wellcome Trust, 2016; British Medical Association, 2015; Clemence et al., 2013). While commercial involvement is still sometimes construed as a “necessary evil” despite these concerns, its necessity is predicated on a clear benefit to a large population (such as for the development of novel pharmaceuticals). Projects that financially benefit private entities without also benefiting larger populations draw particular ire (Grant et al. 2013). As Aitken et al. (2018) characterize a large focus group study on this question, “no one spoke of societal benefits in terms of economic benefit.” In whatever ways we seek to define the public interest, then, it is clear that the public is itself skeptical of private entities as beneficiaries thereof.

However mealy its definition may be, we can also observe that “the public interest” is more than a token phrase, but serves a substantive guiding role within UKB deliberations about data use. For instance, in the midst of public controversy following a hotly contentious study on the “genetic correlates of income,” a reminder email was

sent to all UKB users about appropriate data use which stressed “public interest” as the first and most important criteria (archived in Pitelli, 2019). Elsewhere, UKB documents have probed what might constitute the public interest in health research by considering the ethics of relevant edge cases. In a previous FAQ document on their website (since removed, but quoted in Capps and van der Eijek, 2014), they take up the possibility of collaboration with researchers funded by the tobacco industry but nevertheless pursuing “health-related” projects. This is a historied problem for research ethics: it rehearses the fact that medical research can be used for wide-scale harm by corporate actors even when its factual specifics are defensible, such as how work on the genetics of lung cancer was mobilized by tobacco firms to distract from the carcinogenic contributions of smoking. As the UKB FAQ summarizes,

“Previous research into the effects of smoking saves many millions of lives around the world every year. The UK Biobank Resource is well placed to provide more health information to tackle smoking-related diseases. Researchers using the Resource will have to show that they are bona fide health research scientists and that their work is for the public good. It is virtually impossible to see that an application by the tobacco industry to use the Resource would fulfil these requirements and be approved. Likewise applications by researchers funded by the tobacco industry (directly or indirectly) would be similarly unlikely to be approved.”

This implicitly recognizes that corporate actors may have motivations worth interrogating beyond whether their specific aims might be considered “health-related.”

We do not mean to suggest that insurers, like tobacco companies, are “merchants of doubt” (Oreskes and Conway, 2010) in their production of medical research: their primary goals are certainly not in the manipulation of public opinion. They do, however, have their own troubling incentives. As corporations with fiduciary duties to shareholders, it should not be surprising that an insurer’s financial obligations to increase profit, mitigate loss, and maintain solvency are the primary motivation for business decisions. Insurers are ‘among the most pervasive and powerful institutions in society’ (Ericson et al., 2003: 3): they are gatekeepers that control access to essential services and basic security in people’s lives, and hold a unique social function as a form of ubiquitous biopolitical governance (Lobo-Guerrero, 2011). Yet, despite the responsibilities and expectations that normally come with such important—and such highly-visible—positions in society, insurers are still private corporations that act in all the ways private corporations are impelled to do (Cieply, 2013). We should always keep this basic political economic fact in mind when understanding the motivations and actions of insurers.

Nevertheless, insurers have constructed a variety of discourses linking their market practices to the public interest.

Of particular note here are programs often described under the banner of “behavioral insurance,” where data about an individual consumer’s behavior is used in rate setting: such policies are increasingly available for automotive, health, and life insurance, and capture data through a variety of technologies including wearable devices, vehicle telematics, and smartphone apps (Meyers and Hoyweghen, 2018; Sadowski et al., 2024). Insurers have argued that such programs incentivize positive behavioral changes—like braking less aggressively, or exercising more—through pricing rewards and penalties. By doing so, these so-called “shared value insurance” schemes purport to generate value for both consumers (via reduced rates) and insurers (via reduced losses), binding together imaginations of social and economic value (Sadowski et al., 2024). Such projects mobilize the “moral risks” identified by Ericson et al. (2003) quite literally, assigning financial penalty to “irresponsible” behavior. While we return to trouble the idea that personal data are meaningfully modifiable later in this section, we underline that the invocation of “the public interest” here relies on the idea that certain risk factors and outcomes—and the data that serve as proxies for them—are under a customer’s control, and that insurance pricing can provide a useful incentive to create change for the betterment of society.

Despite their focus here on the behavior of particular consumers, insurers have never limited their interests to data collected directly from individuals. On the contrary, the type of detailed, aggregate data collected by biobanks has the potential to offer significantly more value. From the industry’s point of view, these resources are uniquely capable of providing population-level evidence of risks: these can be used to create classifications, segment populations, underwrite policies, and adjudicate claims—all of which then improves structural stability in the insurance market (Born, 2019). These practices plug into a broader dynamic known as “inverse selection,” in which insurers use technologies like big data and AI analysis to gain an information advantage over customers (Brunnermeier et al., 2023). This asymmetrical relationship allows insurers to screen bad risk and select good risk in evaluating people as insurable assets. Importantly, the dynamic of inverse selection does not have an ultimate endpoint that can be reached once a certain amount of data has been acquired: there is no single event where total risk control has been finally achieved. Rather it is an ongoing process of accumulating as much data as possible, which then feeds into increasingly more powerful (and valuable) forms of risk analysis and risk management (Sadowski, 2023). From this perspective, the longitudinal and population-scale data possessed by national biobanks like UKB represent an essentially unparalleled source of informational value.

Here, we pause to note that it is unclear if the UKB has shared *genetic* data with insurers, or has only provided them with other types of health information. This ambiguity is

itself indicative of a major problem in how these partnerships are negotiated and overseen. While insurance discrimination can be produced using many forms of health data, genetic data are uniquely sensitive. This should necessitate more—not less—transparency about data-sharing from institutions which hold genetic data. Indeed, the prospect that insurers might gain access to genetic data has long been a source of public controversy (Hall et al., 2005; Wauters and Van Hoyweghen, 2016). It is readily apparent that actuarial tools for genetic analysis would result in creating new forms of personalization, categorization, discrimination, and exclusion. Such outcomes raise trenchant justice concerns because they are based on features that are inherent and immutable; for these to inflect the provision or pricing of insurance would represent a particularly galling form of insurance discrimination.

Importantly, concerns about genetic insurance discrimination are not just a paranoid fantasy of academic tech criticism. Surveys of both the general population and of patients with chronic illness have found that most people, regardless of demographic or occupation, are mistrustful of how insurance companies may use genetic data to make decisions (Keogh et al., 2017; Prince et al., 2021). One systematic review of the literature from the United States, Canada, Australia, and Europe found “considerable levels of concern about genetic discrimination” among the public (Wauters and Van Hoyweghen, 2016: 275). Worth underlining here is that these fears drive poor health outcomes *even if insurers aren’t currently using genetic data*. Concern that insurers *might* one day be given the ability to use genetic data in their decision-making is enough to discourage genetic testing amongst the public for fear of producing what could later become derogatory information—even when such testing could allow for improved diagnostics or therapeutics. Such an effect is indicative of a broader moral boundary with clear implications for data governance: genetic data should never be shared with insurers and the details of *any* data-sharing between biobanks and insurers should be transparently disclosed so as to mitigate any concerns of genetic insurance discrimination.

Similar concerns have motivated the establishment of protective regulation: for example, a factor driving the passage of the US Genetic Information Nondiscrimination Act (GINA) “was to encourage greater participation in genetic testing and research by assuaging public fears of genetic discrimination” by insurers (Prince et al. 2021: 343). Importantly, though, GINA only covers the *health* insurance market: insurers are explicitly allowed to use genetic data when determining other sorts of policies, including life and disability insurance. While the regulatory landscape varies globally, few countries offer firmer protections. In Australia, for example, despite a federally mandated partial moratorium, a recent survey of individuals with cancer-predisposing genetic variants found evidence that indicated “both legal (permitted under current regulation)

and illegal discrimination is occurring” by insurers, which had “material impact on consumers” (Tiller et al. 2020: 108, 113).

Although UKB governance is nominally oriented toward UK industry and population, it is important to note that many of its data-sharing agreements are with international actuarial firms. The impacts of how this data affects actuarial risk assessment will play out globally, shaped by the hodgepodge of local legislation and loopholes. Nevertheless, it is worth acknowledging that the UK offers at least slightly more stringent protections. Great Britain has a renewable moratorium based on a formal agreement between the Association of British Insurers (ABI) and the British government, which puts strict limits on how insurers can use predictive genetic testing. The only exception is for high-value insurance policies (Gov UK 2018). Similarly to popular concerns undergirding GINA in the United States, the ABI has noted that the express purposes of this moratorium—and its routine review—is to “provide reassurance to the public that the insurance industry will seek to manage the need for any future change via the [moratorium’s] Code” (ABI 2023). In other words, insurers and regulators alike have sought to assuage public concerns about genetic discrimination through the establishment of governance mechanisms meant to safeguard and steward the public’s interest. We return to this notion of shared responsibility in a moment.

First, though, recall that insurers have described their use of behavioral data to drive price setting as in the public interest, economically incentivizing safer or healthier behavior. Fears of genetic discrimination animate much of the popular discourse because of the obvious immutability of genotype. But the extent to which many lifestyle and behavioral traits are *actually* or *equitably* modifiable is a point worth contending. To give a few examples, here is a very partial list of lifestyle and behavioral traits associated with morbidity based on research with UKB data: unemployment (Pearce et al., 2021); lack of completed education (Davies et al., 2018); lack of nearby green space (Wan et al., 2022); social isolation (Elovainio et al., 2017); not being a “morning person” (Knutson and von Schantz, 2018); being poor, even if you’re in physical shape (Paudel et al., 2023); and living in areas with poor air quality, even if you eat lots of vegetables (Wang et al., 2022). The countervailing variables were explicitly part of the study design—not just controlled traits—in the latter two examples, suggestive of the fact that even “good choices” cannot straightforwardly offset bad personal circumstances.

While all of these examples are ostensibly modifiable behaviors, they are significantly nontrivial and clearly stratify along lines of marginalization—particularly class and race. The opportunity to move to a greener or less polluted area is straightforwardly inaccessible for many; employment and educational status are linked to a variety of factors that may be outside of an individual’s control.

While these studies were conducted by academic researchers, they were all produced using the UKB’s detailed health and behavioral data. That similar findings could be used by insurers to inform their risk models, make pricing decisions, or consider coverage exclusions is self-evident, and yet far exceed individuals consumers’ ability to prevent or remedy (c.f. Ericson et al. on insurance and individual responsibility). As Fourcade and Healy (2013) have also observed, the demarcation of risk categories doesn’t simply *identify* an individual’s life-chances, but works to *determine* them. This is all to say, even without access to more sensitive genetic data, there remains a clear possibility for discriminatory harm—disproportionately allocated to already-marginalized groups—based on the identification of these and similar categories. As a result, insurers’ assertions that categorization practices using large-scale health data may motivate healthy behaviors should be regarded with significant suspicion.

To this point we have contended that the use of UKB data for actuarial research is unlikely to improve health outcomes, and may actually drive worse outcomes while exacerbating extant health inequality. These arguments assume health outcomes are the most important indicator of the public’s interest in health data—a position the UKB itself seems to share based on its assessment of tobacco industry research. Insurers, however, implicitly frame a different understanding of public interest: one which shifts the responsibility for ethical data governance wholly onto regulators, despite their nominal commitments to shared stewardship. One line of reasoning here figures that *denying* insurer access to big data would cause harm to the public. The industry argues that detailed data, at both personal and population levels, is all that ensures the structural stability of the insurance industry. Any uncertainty or constraint in how insurers select risk in the market could lead to a “death spiral” where “harm to both insurers and the public could outweigh any benefit derived from restricting information as premiums could increase across the board,” or otherwise result in mass exclusions from coverage (Prince et al. 2021: 342). In other words, the “right to underwrite” based on publicly-aggregated health data is framed by insurers as an existential necessity for the industry. In this telling, the public interest argument starts to sound more like a threat: *if we go down, we will pull everyone else down with us*.

Elsewhere, insurers have conversely looked to regulators to put limits on their data practices in order to ensure the health of the industry and avoid potential doomsday scenarios caused by data-driven hyper-personalization in the insurance marketplace. As Colm Holmes—then CEO of Aviva and now CEO of Allianz Holdings, two of the world’s largest multinational insurers—said in an interview with a trade magazine: “The use of data is something I think regulators will have to look at, because if you get down to insuring the individual, you don’t have an insurance

industry—you just create people who don't need insurance and people who aren't insurable" (Littlejohns, 2020). This is a remarkable statement: a top executive publicly concerned that insurers will push their own industry over the edge, and looking to regulatory intervention as the solution. It also runs directly counter to the previous set of arguments, leaving us with a strange set of contradictions—detailed population data is required to keep the industry afloat, but is simultaneously poised to sink it. While these imagined futures are at odds, however, they share a common set of first principles: that the health of the insurance industry is definitively in the public interest, and all policy regarding their use of data should begin from this point. Whichever way one feels about the social necessity of insurers, this underlines the fact that any beneficial outcomes for public health that may result from actuarial research using biobank data are incidental to, or in service of, their *financial* interests. As we have already shown, such beneficial outcomes are themselves unlikely, and the public do not share this concern for the financial health of insurers.

From risky individuals to risky groups: De-identification and emergent group interests

Executive Director Collins' second defense of the UKB hinged on the de-identification of the data given to insurers—a common protection for biobank research subjects which involves removing personal identifiers from shared data. Here, Collins' implicit argument seems to be that because de-identified records can't be traced to individual participants, there is no potential for insurance discrimination against an individual based on information they've shared with the UKB, and thus no reason for public concern. This is a potentially dubious claim: while again it's unclear if genetic information was included in material shared with insurers, such data are in essence impossible to truly anonymize (Erlich et al., 2018). It's highly unlikely that commercial insurers would attempt to end-run discrimination laws by re-identifying participants with this data, yet it is also understandable their ability to do so could give participants pause.

But let's take Collins' claim at face value, and accept for the moment that this data is impossible to re-identify. Let's even accept that insurers have no desire to re-identify data at a personal level (though they express that desire regularly). Even under these circumstances, we argue that this move—to use individuated risk as a key evaluative structure in determining whether research is in the public interest—is itself a significant ethical lapse. This reasoning innately assumes that the public is made up of individuals who only have claim to defend or advance *self*-interest, rather than group or external interests. As a result, when no

personalized risk is involved, this logic then positions the UKB as a paternalistic arbiter in the equitable distribution of benefits. In the previous section we have already questioned whether insurance research can actually be understood to offer social benefit in the public interest in any real sense. Here we examine how the rhetorical deployment of de-identification works to further degrade what role the public can play in this debate. We also contrast alternative frameworks for health data governance that are predicated on the articulation of group, rather than individual, interests.

The tension between individual and group interests is not unique to the UKB: it is inherited, in part, from the Nuffield Council on Bioethics—a leading ethics watchdog in the UK whose work informs UKB policymaking. On one hand, Nuffield guidance articulates that emerging technologies have the ability "to affect social relations and to shape the conditions of common life in non-trivial ways, potentially changing the future options available to all in ways that may favour only some" (NCoB, 2012: XX). This would seem to position the societal-level harms of insurance technology as clear grounds for intervention. At the same time, the specific protections Nuffield affords are near-invariably granted to individuals. As Sarah Cheung (2020: 8) has argued, "in construing of discrimination as occurring from one-off events, as the Nuffield report does, this diminishes the recognition of substantial cumulative impacts arising from individual 'legitimate' uses of profiling." This is a critical point: much of the potential for harm discussed in the last section is based on how such "legitimate" profiling is likely to affect insurance pricing and allocation in ways which would affect *groups*—groups made up of UKB research participants and nonparticipants alike. The mechanisms and implications of this bear drawing out.

In the previous section we discussed a number of traits—both biological and sociological—that are linked to morbidity: it is easy to imagine how these might become inbuilt factors in a black-boxed actuarial algorithm. But let's now look at a much more clear-cut example, linking a single piece of consumer data to a specific disease. A recent study published in *Nature Medicine* claims to have developed a machine learning system trained on motion data from Apple Watches that can assess a person's risk for Parkinson's Disease up to seven years prior to clinical diagnosis (Schalkamp et al. 2023). (Worth noting: the study additionally uses health records provided by the UKB.) While this project positions itself within a larger body of work involving Apple and Parkinson's research—which has already resulted in three Apple Watch apps being approved by the US Food and Drug Administration to track Parkinson's symptoms (Aguilar 2023)—this new tool is not descriptive, but *predictive*. It assesses disease risk before disease symptoms are evident, much like an insurer would seek to do.

In addition to this type of research partnership, the Apple Watch also has a long partnership with Vitality, a leading

insurance technology platform that works with major insurers in every global region—in addition to also being an insurance provider in the UK and South Africa—to deliver life/health insurance programs based largely on behavioral data collected from customers’ wearable devices (Sadowski et al. 2024). We can now easily imagine how a predictive model based on legitimately acquired customer data from Apple Watches could be used by insurers to identify people at an “early risk for Parkinson’s” long before they show any symptoms or have a diagnosis. From an insurer’s perspective, it would be wasteful to not extract more value from data they already possess. From a consumer’s perspective, it might come as a shock—and an economic hardship, following increased premiums or difficulty getting policies—to learn that they are now classified as higher risk.

While these sorts of direct inferences are one set of possible data harms, actuarial algorithms are typically more subtle in their approach to the delineation of risk. A person may end up identified as belonging to any number of unfavorably “risky” categories by such an algorithm; if they are, they’re unlikely to know how or why they have been categorized. It is generally impossible to draw a straight causal line from a specific data point to a specific outcome for a specific person in the types of complex risk models used by insurers. While we do know for certain that insurers use health data to create new categories of risk for classifying and discriminating against groups as an ongoing process, the specificities of these systems are hidden inside institutional and technological black boxes. This means that health data about Person X could contribute to creating a predictive risk model that is used to sort, segment, and select or exclude Population Z in the marketplace for insurance. Person X may or may not be part of Population Z; they may not know of their relationship to Population Z. Moreover, Population Z may not even be a clearly demarcated group from their perspective—as we have discussed, the kinds of factors that inform risk models are often nonobvious, and the people who share them may appear to have little of importance in common. Nevertheless, data about Person X has now contributed to Population Z’s (in)ability to procure insurance.

The argument made by institutions like the UKB and Nuffield is that Person X only has an interest in their health data if that data is used to directly identify, classify, and discriminate against Person X themselves—and only if the possibility of such use can be proven beyond a doubt. This argument works well in the sterile field of ethics abstracted from society at large: here, there is only a single data source (UKB), a single data user (Insurance Inc.), a single data model (Total Risk), and single impacts (Decision Points) on individual people (Person X). However, this way of thinking about data relations makes a common—but fatal—mistake in how we understand data governance in a complex digital society: it is premised

on what Salomé Viljoen (2021) calls “vertical relations,” in which different actors in the data chain have direct connections from source to collector to user to outcome. By this framing, it appears possible to map, understand, and predict the entire data network and its consequences.

However, the systems we are talking about here are “horizontal relations,” which exist not at individual levels but at population scales (Viljoen 2021). They are flows of data that can link countless people together across many networks over large distances of space/time; a vast multiplicity of sources, collectors, users, models, and consequences of data get mixed up in ways that are impossible to trace if we are still thinking in vertical terms. Importantly, this horizontal way of thinking is not foreign to insurers—actuarial science has always taken the relations within and between populations as their object of analysis and management. Thus it is particularly disingenuous for biobanks and insurers to treat data technologies as horizontal systems while limiting their ethical imagination to vertical logics. Understanding the horizontal relationships at play—between and exceeding UKB participants—demands a much more expansive ethical account.

Framing any analysis of data harm at the level of individuals and events will always miss how the systems work and why they matter; it misses what interests populations, not just people, have in the governance of social data. This is all to say, we can’t build biobank data ethics out of reactive and narrow imaginations of personalized data harms—a more expansive horizontal view is necessary. Data can be harmful even when decoupled from personal identity, and the people and communities which share data about themselves should have a stake in shaping how they’re used even when their data are de-identified.

It is worth briefly considering how the relationship between group harms, benefits, and health data governance has been constructed differently in other cases. While these projects differ sharply in their material histories (and we certainly do not mean to suggest their solutions are straightforwardly portable to UKB policy), they are useful in highlighting the matters of concern that buckle between individuals, groups, and particular mobilizations of “the public interest.”

Working in the aftermath of centuries of expropriative and exploitative data collection, recent scholarship under the banner of “Indigenous data sovereignty” asserts that Indigenous communities should be the owners and stewards of data about themselves—particularly biological and health data (R. Tsosie, 2019; K. Tsosie, 2021; Carroll et al., 2020). This recognizes a history of data practices that have produced harmful claims about Indigenous peoples, as well as a tradition of research “in the public interest” that yet has rarely served the interests of Indigenous communities. In these projects, Indigenous ownership of data—including control over subsequent uses and interpretations as well as claim over benefits—is

the necessary repair. This ethos has been formalized into governance policy by groups like the Native BioData Consortium, a biobank that consolidates both human and nonhuman genetic data. In a different case, Molldrem and Smith (2020, 2023) have problematized data management and the molecular surveillance of HIV as a public health project. While extensive data collection to identify HIV transmission chains is often construed as a public benefit in policy discussions, people living with HIV have argued that the evidence of benefit to broader publics is overstated while risks and harms to their communities have been underemphasized. As Molldrem and Smith argue, novel models of ongoing/recurrent consent, opt-out opportunities, and plain language summaries of data use (at a much more granular level than “for public health interventions”) are necessary, and question the type of broad consent strategies used by most biobanks.

What we learn from contestations of Indigenous data and HIV surveillance, then, is that “the public interest” has often been constructed in ways which not only run against the interests and needs of various groups, but are simultaneously used to justify the extraction of their data to benefit others. These projects emphasize continued reevaluation of policy and community engagement about what constitutes acceptable data use and equitable distribution not only of harms, but also of benefits.

These examples differ significantly from that of the UKB, which does not focus solely on recruitment from marginalized groups, but rather, a “representative population.” Nevertheless, insurance technologies like risk assessment algorithms exist solely to identify novel groupings which can then be discriminated against in the marketplace. Just because these groups have not faced historical marginalization—or, indeed, been previously recognizable as groups—does not mean they share no interest in curtailing insurer access to population data which will be used to construct their market prospects. Put another way, we are all potential risky subjects under the gaze of the algorithm. As legal scholar Tom Baker (2003: 275) notes in reference to the moral economy of insurance, “While some ‘low risk’ individuals may believe that they are benefited by risk classification, any particular individual is only one technological innovation away from losing his or her privileged status.” Incorporating different risk factors into an assessment, using alternative sources of data for analysis, or tweaking the parameters of a model may lead to new ways of creating groups and categorizing people that upturn the existing status quo of risk classification. The implications of this fact bear ethical scrutiny in the construction of data use policies.

Conclusion

Writing about a previous health data-sharing scandal in the UK, Carter et al. (2015) argue for a renewed attention to the

“social license” in ethical data governance. As they describe, compliance with the letter of the law alone is insufficient to ensure public trust: rigorous and ongoing public consultation toward the co-development of norms is a necessary step to retain it, particularly when emerging technologies with uncertain social impact are involved. Here, we turn to observe that Collins’ letter hangs largely on the issue of legal compliance: issues of transparency, consent, and notification processes make up most of his objections to *The Observer* reporting on the UKB’s data practices. *The Observer* reporting, however, did not primarily hinge on regulatory compliance. What it identified was a *breach of trust*.

In conclusion, we argue that this mistrust—coupled with the well-documented public aversion toward the commercial use of health data and imminent possibility of discriminatory harm—demands significantly expanded public consultation and oversight on any data-sharing between biobanks and insurers. This improved governance is the bare minimum when dealing with forms of data-sharing that butt against the moral limits of permissibility. This is also in keeping with Nuffield recommendations, which advocate for participants to be involved “as collaborators in the whole system” of technological development using health records as a matter of “respect for them as persons who have morally significant interests” over the use of their data (p. 91). Such consultation must also be based on a significantly expanded horizon of data governance that accounts for the relational realities of how data is produced and used, especially in complex systems like artificial intelligence and actuarial modeling. The public cannot and should not be talked out of their mistrust by paternalistic promises that de-identification will keep them safe, or that the development of commercial insurance technologies will somehow offer social benefit. Instead, this moment of outrage should be understood to spotlight a violation of the social license which will necessitate careful effort toward repair. It should also be an opportunity to reconstruct a regime of data governance that goes beyond merely assuaging public concerns, but instead is fit for the purpose of securing the public interest against the hazards of private desires. We contend that this will necessitate sharply restricting and closely reviewing—if not outright prohibiting—insurers’ access to biobank data.

It’s worth underlining that this issue is not unique to the UKB, nor to the UK. The types of health prediction algorithms produced through such actuarial research will undoubtedly be used to make crucial decisions about those with nebulous health and behavioral “risk factors” broadly, no matter where their training data originate. Moreover, similarly large collections of health data proliferate globally, and their access policies are often as or more permissive than the UKB’s. Other public organizations like the massive All of Us Project (the US equivalent to UKB) do not currently prohibit actuarial firms from

applying for access: while it's unclear if such applications would be approved, the lack of clear guidelines is certainly cause for concern. More troublingly, large commercial databases have a clear economic incentive to share their data with insurers for the right price and offer even less transparency for their users. Many 23andMe users, for instance, are not aware that "consent for research" includes consent for their data to be shared with *corporate* researchers. When asked about this in a recent interview, CEO Ann Wojcicki demurred that "it's not individual-level data" that is being shared (Germain 2024). It's interesting to observe how often the rhetoric of de-identification is used to abstract data donors from a stake in how their data travel.^[2] It's a dismissive gesture, as if to say *stop worrying about it, such things don't directly concern you.*

The UKB incident should spur urgent attention to the broader bioethics of actuarial research—a badly underdeveloped thread of scholarship, particularly given the central role insurers play in determining health outcomes through the uneven provision of resources. Simultaneously, it demands we reconsider what we understand the public's interest in collective health data to be. Is it, as the UKB contends, simply to be protected from individually harmful misuse? Or can we understand data governance as a more expansive site of negotiation over what a desirable future looks like, and who has the authority to shape it?

Acknowledgements

The authors would like to thank the reviewers for their thoughtful comments on this project. They would also like to thank Chris O'Neill for introducing them, and sparking a happy collaboration that wouldn't have happened otherwise.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Kathryne Metcalf's work was supported by the National Science Foundation (grant number 2341622). Jathan Sadowski's work was supported by the Australian Research Council (grant number DE220100417).

ORCID iDs

Kathryne Metcalf  <https://orcid.org/0000-0001-6529-1695>
Jathan Sadowski  <https://orcid.org/0000-0002-0324-708X>

Notes

1. Not every type of data is collected about every participant, but in general the data collection process is extensive and often involves recontacting participants at later dates to collect even more information.

2. This is not the first time 23andMe has suggested that "de-identification" obviates the need for conventional ethical considerations. In an early scandal, the company put forward a study based on data for which neither IRB review nor traditional informed consent had been attained. They argued that the project did not constitute human subject research because researchers only had access to de-identified data, posing the company as an anonymizing intermediary and data broker rather than a responsible party to the research—despite the fact that authors on the manuscript were also employed by 23andMe. While the study was eventually published in *PLoS*, journal editors complained that 23andMe's position was "preposterous" and took advantage of an "unfortunate loophole." For a more thorough account of this incident, see Spector-Bagdady (2016).

References

ABI (2023) Code on Genetic Testing and Insurance. Accessed 22 February 2024: <https://www.abi.org.uk/data-and-resources/tools-and-resources/genetics/code-on-genetic-testing-and-insurance/>.

Aguilar M (2023) For Parkinson's disease, advances spurred by Apple Watch offer a glimmer of hope. *Stat News*, 24 October. Accessed 9 February 2024: <https://www.statnews.com/2023/10/24/apple-watch-parkinsons-disease-apps-parky-strivepd-neurorpm/>.

Aitken M, Porteous C, Creamer E, et al. (2018) Who benefits and how? Public expectations of public benefits from data-intensive health research. *Big Data & Society* 5(2): 205395171881672.

Baker T (2003) Containing the promise of insurance: Adverse selection and risk classification. In: Ericson RV and Doyle A (eds) *Risk and Morality*. Toronto: University of Toronto Press, 258–283.

Bick AG, Metcalf GA, Mayo KR, et al. (2024) Genomic data in the all of Us research program. *Nature* 627: 1–7.

Birch K and Tyfield D (2013) Theorizing the bioeconomy: Biovalue, biocapital, bioeconomics or . . . what? *Science, Technology, & Human Values* 38(3): 299–327.

Born P (2019) Genetic testing in underwriting: Implications for life insurance markets. *Journal of Insurance Regulation* 38(5): 1–18.

British Medical Association (BMA) (2015) Secondary Uses of Healthcare Data Public Workshop Debrief. Available at: <https://www.bma.org.uk/collective-voice/policy-and-research/ethics/secondary-uses-of-data/public-workshop>.

Brunnermeier MK, Lamba R and Segura-Rodriguez C (2023) Inverse Selection. *SSRN*, 5 December. Accessed 1 February 2023: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3584331.

Bycroft C, Freeman C, Petkova D, et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562(7726): 203–209.

Capps B (2012) The public interest, public goods, and third-party access to UK Biobank. *Public Health Ethics* 5(3): 240–251.

Capps B (2013) Defining variables of access to UK Biobank: The public interest and the public good. *Law, Innovation and Technology* 5(1): 113–139.

Capps B and van der Eijk Y (2014) The tobacco industry, researchers, and ethical access to UK Biobank: Using the public interest and public good. *American Journal of Public Health* 104(10): 1833–1839.

Carroll SR, Garba I, Figueroa-Rodríguez OL, et al. (2020) The CARE principles for indigenous data governance. *Data Science Journal* 19: 43–43.

Carter P, Laurie GT and Dixon-Woods M (2015) The social licence for research: Why care.data ran into trouble. *Journal of Medical Ethics* 41(5): 404–409.

Cheung S (2020) Disambiguating the benefits and risks from public health data in the digital economy. *Big Data & Society* 7(1): 205395172093392.

Ciepley D (2013) Beyond public and private: Toward a political theory of the corporation. *American Political Science Review* 107(1): 139–158.

Clemence M, Gilby N, Shah J, et al. (2013) Wellcome Trust Monitor Report Wave 2: Tracking Public Views on Science, Research and Science Education. Available at: <https://wellcome.org/sites/default/files/monitor-wave2-full-wellcomemay13.pdf>.

Cochran CE (1974) Political science and “the public interest.”. *The Journal of Politics* 36(2): 327–355.

Collins R (2023) Response to highly misleading article in The Observer. Nov 12. Available at: <https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/response-to-highly-misleading-article-in-the-observer>.

Croteau D and Hoynes W (2006) *The Business of Media: Corporate Media and the Public Interest*. Newbury Park: Pine Forge Press.

Dahl RA and Lindblom CE (1954) *Politics, Economics, and Welfare*. London: Routledge.

Das S (2023) Private UK health data donated for medical research shared with insurance companies. Nov 12. *The Observer*. <https://www.theguardian.com/technology/2023/nov/12/private-uk-health-data-donated-medical-research-shared-insurance-companies>.

Davies NM, Dickson M, Davey Smith G, et al. (2018) The causal effects of education on health outcomes in the UK Biobank. *Nature Human Behaviour* 2(2): 117–125.

Douglass B (1980) The common good and the public interest. *Political Theory* 8(1): 103–117.

Downs A (1962) The public interest: Its meaning in a democracy. *Social Research* 29(1): 1–36.

Elovainio M, Hakulinen C, Pulkki-Råback L, et al. (2017) Contribution of risk factors to excess mortality in isolated and lonely individuals: An analysis of data from the UK Biobank cohort study. *The Lancet Public Health* 2(6): e260–e266.

Ericson R, Doyle A and Barry D (2003) *Insurance as Governance*. Toronto: University of Toronto Press.

Erlich Y, Williams JB, Glazer D, et al. (2014) Redefining genomic privacy: Trust and empowerment. *PLOS Biology* 12(11): e1001983.

Fortun M (2008) *Promising Genomics: Iceland and deCODE Genetics in a World of Speculation*. Oakland: University of California Press.

Fourcade M and Healy K (2013) Classification situations: Life-chances in the neoliberal era. *Accounting, Organizations and Society* 38(8): 559–572.

Galston WA (2007) An old debate renewed: The politics of the public interest. *Daedalus* 136(4): 10–19.

Germain T (2024) 23andMe wants to double down on DNA data mining with its business in a tailspin. Feb 14. *Gizmodo*. <https://gizmodo.com/23andme-admits-mining-your-dna-data-is-its-last-hope-1851252582>.

Gov UK (2018) Code on Genetic Testing and Insurance. Accessed 7 February 2024: <https://www.gov.uk/government/publications/code-on-genetic-testing-and-insurance>.

Grant A, Ure J, Nicolson DJ, et al. (2013) Acceptability and perceived barriers and facilitators to creating a national research register to enable ‘direct to patient’ enrolment into research: The Scottish health research register (SHARE). *BMC Health Services Research* 13(1): 422.

Hall MA, McEwen JE, Barton JC, et al. (2005) Concerns in a primary care population about genetic discrimination by insurers. *Genetics in Medicine* 7(5): 311–316.

Hartman T, Kennedy H, Steedman R, et al. (2020) Public perceptions of good data management: Findings from a UK-based survey. *Big Data & Society* 7(1): 205395172093561.

Helmreich S (2008) Species of biocapital. *Science as Culture* 17(4): 463–478.

Keogh LA, Niven H, Rutstein A, et al. (2017) Choosing not to undergo predictive genetic testing for hereditary colorectal cancer syndromes: Expanding our understanding of decliners and declining. *Journal of Behavioral Medicine* 40(4): 583–594.

Knutson KL and von Schantz M (2018) Associations between chronotype, morbidity and mortality in the UK Biobank cohort. *Chronobiology International* 35(8): 1045–1053.

Leonelli S (2016) *Data-Centric Biology: A Philosophical Study*. Chicago: University of Chicago Press.

Lippmann W (1955) *The Public Philosophy*. London: Routledge.

Littlejohns P (2020) Aviva’s Colm Holmes on claims inflation, stupid drivers and commercial insurance. *NS Insurance*, 7 February. Accessed 2 February 2024: <https://www.nsinsurance.com/analysis/aviva-colm-holmes-claims-inflation/>.

Littlejohns TJ, Sudlow C, Allen NE, et al. (2019) UK Biobank: Opportunities for cardiovascular research. *European Heart Journal* 40(14): 1158–1166.

Lobo-Guerrero L (2011) *Insuring Security: Biopolitics, Security and Risk*. London: Routledge.

Mahoney JT, McGahan AM and Pitelis CN (2009) Perspective—The interdependence of private and public interests. *Organization Science* 20(6): 1034–1052.

Metcalf K (2022) Affective economies in blood banks and biobanks: Vital accounting from US transfusion medicine to genetic research, 1935–1990. *Social History of Medicine* 35(2): 369–388.

Meyers G and Hoyweghen IV (2018) Enacting actuarial fairness in insurance: From fair discrimination to behaviour-based fairness. *Science as Culture* 27(4): 413–438.

Mitchell R and Waldby C (2006) *Tissue Economies: Blood, Organs, and Cell Lines in Late Capitalism*. Durham: Duke University Press.

Molldrem S and Smith AKJ (2020) Reassessing the ethics of molecular HIV surveillance in the era of cluster detection and response: Toward HIV data justice. *The American Journal of Bioethics* 20(10): 10–23.

Molldrem S and Smith AKJ (2023) Health policy counterpublics: Enacting collective resistances to US molecular HIV surveillance and cluster detection and response programs. *Social Studies of Science* 54(3): 451–477.

Nuffield Council on Bioethics (NCoB) (2012) *Emerging Biotechnologies: Technology, Choice and the Public Good*. London: Nuffield Council on Bioethics.

O'Donoghue S, Dee S, Byrne JA, et al. (2022) How many health research biobanks are there? *Biopreservation and Biobanking* 20(3): 224–228.

Oreskes N and Conway EM (2010) *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming*. London: Bloomsbury Publishing.

Paudel S, Ahmadi M, Phongsavan P, et al. (2023) Do associations of physical activity and sedentary behaviour with cardiovascular disease and mortality differ across socioeconomic groups? A prospective analysis of device-measured and self-reported UK Biobank data. *British Journal of Sports Medicine* 57(14): 921–929.

Pearce M, Strain T, Wijndaele K, et al. (2021) Is occupational physical activity associated with mortality in UK Biobank? *International Journal of Behavioral Nutrition and Physical Activity* 18(1): 102.

Pitelli S (2019) My Letter to the UK Biobank. *Unwashed Genes*, Dec 19. Available at: <https://unwashedgenes.blogspot.com/2019/12/my-letter-to-uk-biobank.html>.

Prince AER, Uhlmann WR, Suter SM, et al. (2021) Genetic testing and insurance implications: Surveying the US general population about discrimination concerns and knowledge of the genetic information nondiscrimination Act (GINA). *Risk Management and Insurance Review* 24: 341–365.

Rajan KS (2006) *Biocapital: The Constitution of Postgenomic Life*. Duke: Duke University Press.

Rawls J (1971) *A Theory of Justice*. Cambridge: Harvard University Press.

Sadowski J (2023) Total life insurance: Logics of anticipatory control and actuarial governance in insurance technology. *Social Studies of Science* 54: 231–256.

Sadowski J, Lewis K and Bednarz Z (2024) Risk, value, vitality: The moral economy of a global behavioral insurance platform. *Economy and Society* 53: 227–249.

Schalkamp AK, Peall KJ, Harrison NA, et al. (2023) Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. *Nature Medicine* 29: 2048–2056.

Schubert GA (1958) The theory of the public interest. *Political Research, Organization and Design* 1(5): 34–36.

Slokenberga S (2021) Setting the foundations: Individual rights, public interest, scientific research and biobanking. In: Slokenberga S, Tzortzatou O and Reichel J (eds) *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation Across Europe*. New York: Springer International Publishing, 11–30.

Sorauf FJ (1957) The public interest reconsidered. *The Journal of Politics* 19(4): 616–639.

Spector-Bagdady K (2016) “The Google of healthcare”: Enabling the privatization of genetic bio/databanking. *Annals of Epidemiology* 26(7): 515–519.

Stevens R (1998) *American Medicine and the Public Interest*. Oakland: University of California Press.

Strasser BJ (2019) *Collecting Experiments: Making Big Data Biology*. Chicago: University of Chicago Press.

Tiller J, Morris S and Rice T, et al. (2020) Genetic discrimination by Australian insurance companies: a survey of consumer experiences. *European Journal of Human Genetics* 28: 108–113.

Tsosie KS (2021) Inclusion without equity: the need to empower Indigenous data sovereignty in precision health. In: Lorusso L and Winther RG (eds) *Remapping Race in a Global Context*. London: Routledge, 148–154.

Tsosie R (2019) Tribal data governance and informational privacy: Constructing indigenous data sovereignty. *Montana Law Review* 80(2): 229–268.

UK Biobank Access Policy version 2.1 (2022, July) <https://www.ukbiobank.ac.uk/media/hfnf2w0f/20220722-access-procedures-v2-1-final.pdf>.

UK Biobank Ethics and Governance Framework version 3 (2007) <https://www.ukbiobank.ac.uk/media/0xsbmfmw/egf.pdf>.

Viljoen S (2021) A Relational Theory of Data Governance. *Yale Law Journal* 131(2): 573–654.

Wan S, Rojas-Rueda D, Pretty J, et al. (2022) Greenspace and mortality in the U.K. Biobank: Longitudinal cohort analysis of socio-economic, environmental, and biomarker pathways. *SSM – Population Health* 19: 101194.

Wang M, Zhou T, Song Q, et al. (2022) Ambient air pollution, healthy diet and vegetable intakes, and mortality: A prospective UK Biobank study. *International Journal of Epidemiology* 51(4): 1243–1253.

Wauters A and Van Hoyweghen I (2016) Global trends on fears and concerns of genetic discrimination: A systematic literature review. *Journal of Human Genetics* 61(4): 275–282.

Wellcome Trust (2016) The One-Way Mirror: Public attitudes to commercial access to health data. Available at: <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf> (accessed 13 September 2019).