

RAPID ETHICS REVIEW – UK Pandemic Ethics Accelerator

The ethics of data self-reporting: a review of important issues and best practices

Part I: ethical issues

Niccolò Tempini

University of Exeter & Alan Turing Institute

n.tempini@exeter.ac.uk

This report commissioned by UK Pandemic Ethics Accelerator. Revision submitted 24th June 2022. Please cite as follows:

Tempini, N. 2022. The ethics of data self-reporting: a review of important issues and best practices – Part I: ethical issues. UK Pandemic Ethics Accelerator (under review). Preprint available at <http://dx.doi.org/10.5281/zenodo.7324049>

Headline points

- The spectrum of self-reporting is extremely diverse and impossible to govern through a one size fits all approach; projects can be scientific or not, can raise important privacy concerns, and can be open to organised manipulation and poor governance of risks and harms; best practices and guidelines must be adapted to the local setting and remain open for improvement
- People take part in data collection for the most heterogeneous of purposes. Closely controlling hopes, aims and beliefs of participants is beyond the reach of any sensible self-reporting exercise; this heterogeneity introduces various biases in the data, and adds unpredictability to the data collection project
- Governance of self-report data collection is a thorny issue; new data governance approaches have been put forth to deal with the heterogeneity of ends that big data collection projects now need to be able to support, that require some experimentation, but the best chance of long-term sustainable governance comes from collective governance frameworks

- Greatest scientific potential of self-reporting is through data linkage with other forms of heterogeneous data, however these forms of data linkage are also the source of the highest risks for privacy and individual harms; flexible and sustained ethical oversight is key
- Participation in self-reporting is not equally distributed across society; when there are benefits associated in participation, it can exacerbate existing inequalities

The ethics of data self-reporting

Part I: ethical issues

1. Introduction

As the first wave of the c19 pandemic washed through Europe in early 2020, governments desperate to get a grip and researchers and companies hoping to make direct impact into the issue all identified the use of self-report data (data that are collected by participants) as one of the key weapons digitalised societies have to make sense of a health phenomenon that is poorly understood, distributed and urgently concerning.

Data had to be collected for at least **a few different purposes**, including symptom tracking and illness management; contact tracing (both at the national and private levels); and test result reporting. These purposes could be met in various combinations by many different offerings, and governments endorsed a spectrum of different technical methods for the collection, and legal requirements and implications tied to the use or non-use of these apps (Kitchin 2020). Data also had to be collected through **different levels of direct involvement** on the part of the participant. Some of the data collection is automatically performed by the app once the participant has set it up; other data can only be produced if the user responds to specific questions.

The enthusiastic ‘technological solutionism’ (Morozov 2014) with which some governments, including the UK, brandished a mobile application as a godsend for keeping the virus at bay can be levelled against their desperation to make their promises believable, the sacrifices

digestible, and keep a stable social order at a time when the sky was falling. While uncritical *'there's an app for that'* arguments are easily dismissed, the more modest hypothesis, that self-report data can be a valuable resource in specific circumstances, is not. Data self-reporting for the study of health phenomena is a strategy clinical researchers have been using for decades and well before the rise of Internet and personal computing, patients were asked to self-report symptoms and experience through paper-based questionnaires. The spectrum of patient experience is of course enormously varied so it goes without saying that the same approach will yield highly reliable results with self-reporting about one medical condition, while patients suffering from another condition might not be able to report good quality data.

The purposes for citizens to become a data contributor can be many; the general (im)mobilisation that a pandemic can bring to a population is an extraordinary force for project enrolment. Contact tracing apps risk reinforce social injustices, better protecting those who are already better off, while interfering with the livelihoods of more precarious, frontline workers (Ada Lovelace Institute and The Health Foundation 2021). It is inadequate, as Lucivero and colleagues (Lucivero et al. 2021) remind us, that contact tracing apps are imposed on citizens on the grounds of a false dichotomy between individual privacy and public safety (Lucivero et al. 2021).

Outside of pandemic times, data self-reporting has often been associated with individual and collective patient empowerment, and patient movements. In 'normal' time, it has long been clear that patients who are able to engage on large scale, distributed data collection exercises are better able to advance demands and push for a reorganisation of the research and policy agenda concerning their condition (Epstein 1996; Rabeharisoa, Moreira, and Akrich 2014). Data self-reporting becomes a way, then, to put things into motion 'from the bottom up.'

However, data self-reporting is also the basis on which tech giants have been building empires of surveillance now caught in the public eye for their potential to manipulate and interfere with the lives of individuals, communities and countries. A great deal of criticism has been levied towards project and rhetoric of web-based participation originating in the environs of the global tech industry, and disclosure of personal information, traits and behaviour through automated or manual self-reporting is seen by many as a driver of disempowerment and a new colonialism (Zuboff 2019; H. Ekbja and Nardi 2014). The pandemic has created opportunities for corporate technology platforms to further penetrate

public health systems (Lucivero et al. 2020; Tempini 2022), e.g., the co-development by Google and Apple of contact tracing methods to be used by the NHS in its contact tracing app; leading to calls of ‘covidwashing’ (Kitchin 2020); worries of data function creep were the platform monopolists able to directly or indirectly make use of the data for private initiatives (Lanzing 2021); and ultimately, the contribution to a general trend were collective dependence from the proprietary technology platforms of monopolists deprives the public from the ability to imagine and develop a future without it (Sharon 2020).

This report relates to both the exceptional pandemic times and the dynamics of accelerated technology adoption and intensified data collection that they have generated; and the ‘state of play’ in normal times, when individual have been involved in voluntary and involuntary contribution of data by a range of projects from consumer grade e-commerce technologies to ‘cognitive surplus’ (Shirky 2010) collective projects. In highly digitalised societies, the topic of the ethics and governance of self-report data is inevitably inexhaustible. Given the space restrictions, the discussion offered in this report needs to be limited to the areas and problems that are perceived as important for the most.

The report is divided in two parts. This document, *Part I: ethical issues*, outlines the ethical issues that are outstanding the pandemic experience and the decades preceding it in which web technology became mainstream. A second document, *Part II: best practice innovations*, takes stocks of innovations that can help improve best practices with a view to tackle challenges outlines in Part I.

2. Ethical issues

2.1 Function creep

Data have a tendency to function creep: they often go on to be reused for many other purposes than the ones initially envisaged. The technology used to organise data collection tends to function creep too. This is true both in private and public sector projects. It is worrying because if data are not risk-managed in the best way, disclosures and leaks that lead to individual and group harm are more likely.

In the private sector, the emergence of *surveillance capitalism* (Zuboff 2019), and the lively debate discussing it, has highlighted the deceitful and pernicious intrusions of privacy and the

subsequent behavioural manipulations that US tech giants have built their empires from: an almost two decades long thread of surreptitious projects to take ever more and diverse data about users; flanked by an assertive and crafted rhetoric on the value of sharing, emancipation, community, and peer2peer entrepreneurship; that increased the ability of these giants to predict, modify and generate human behaviour to the point that legitimate concerns have been raised as to the effects of big tech services on just about anything social including politics, markets and mental health. The cost of surveillance capitalism, as Zuboff observes, is *human futures*. The model has been imitated by companies throughout the sector to the point that both practitioners and public, in different ways, are failing to imagine how tech can be built without relying on the sale of user profiles and behavioural modification. Innovative organisations focusing on privacy-first products are struggling to be seen and break through the current market chokehold. For this reason, the offer by Google and Apple to collaborate in developing an interoperable framework for contact tracing met mixed reception. In a time of great emergency it was congenial for national covid response efforts to develop their apps on top of this tech giant-concerted framework. The UK tried to independently develop the same functionality for its NHS contact tracking app, only to backtrack. The tech giants were praised for munificence and sense of civic responsibility. The framework also employs gold standard privacy-preserving techniques approved by privacy experts. It was easy to see, however, how self-serving the move could be to white-wash the giants' reputations with the very same broad framework they had been tarnished with – individual-level, comprehensive digital surveillance (Kitchin 2020). The offer to help fits a history of manipulative attitudes to the public discourse.

The legitimisation of surveillance technology, methods and frameworks as a response to the pandemic emergency has been such that even organisations who have been known to the public for nothing but scandals have had a chance at 'covidwashing'. For instance, Israel's contact-tracking app was developed by NSO Group, a secretive spyware organisation notorious for its services to authoritarian governments across the world (Kitchin 2020). Similarly, Palantir and Experian, other actors of the digital economy of questionable ethics reputation, have also taken part. Observing how easily techniques from the most controversial commercial surveillance practices can be transferred all the way to state infrastructure for pandemic response should make us ask what we are dealing with. For companies such as Palantir, the opportunity was not only reputational, but also an opportunity to turn private technology into state infrastructure, with the commercial benefits that can ensue in the long

run (Tempini, 2022). The means are the translation, adaptation and re-deployment of technology, methods and resources used in population surveillance. The aim is to become indispensable infrastructure of public health response and prepare state administration processes for further penetration through new drives of technologically, and organisationally, compatible systems provision. Case in point, after providing an analytics infrastructure crucial to the coordination of the covid response, Palantir now looks set to win an unprecedentedly valuable and wide ranging contract for the provision of future NHS digital capabilities (*Financial Times* 2022) extending from the infrastructure already provided during the pandemic. The strategy is right out of the infrastructure studies playbook.

Despite the scandals revealed by a number of whistleblowers and leaks (e.g. Snowden on the PRISM program) reaction over recent years has been mixed and, so far, shy of introducing game-changing protections; such is the attraction that the public sector feels in respect to these new sources. The single-most impactful law recently issued on the topic, the EU(UK) GDPR, has had an uncertain track record as a sensitivising device, as broad patterns of use have not shifted. It provided great flexibility in implementation. In it, the opportunity for technology providers' is implementation fragmentation. If every provider implements the regulation in a slightly different way (as in the myriad different ways to do 'cookie' language, explanation, and notification interface), user burden in expressing and articulating privacy preferences is highest. It is possible that it is rather a string of publicised scandals, such as the Facebook/Cambridge Analytica one, involving the threat of election manipulation and disinformation warfare, that has finally started to bite.

Internet researchers, in the Internet and beyond, should also question and account for the ways in which they might be benefiting from such widely criticised ways to generate data, and the ways in which their own practice can relate to function creep issues: "*Similarly, harvesting sensitive information from public-private environments such as social media may raise ethical issues, especially for research involving vulnerable populations who may have limited understanding of the implications of disclosing personal information on these platforms*" (Clark et al. 2019).

In the public sector, large trends in networking and ubiquitous computing have brought about an explosion of digital watching as made possible by a highly granular fabric of recording devices; and a network of interoperable databases that are increasingly interconnected for

fast, increasingly real-time, networked access. From CCTV cameras, which the UK has led the way in ubiquitous deployment and acceptance, to a panoply of digital listening devices, huge amounts of data can be made available to various agencies from the law enforcement to public administration. The value of data and the pressures to fuel the growth of a competitive digital economy have created incentives for governments to enable the regulated access to previously unavailable data even down to the individual level. The unveiling in Spring 2021 of plans by the UK Government for the regulated access and re-use of NHS England and Wales data (see GDPR reporting; (Machirori and Patel 2021)) is the latest of a series of attempts to inject these data in the digital and research sector economies (Vezyridis and Timmons 2017) and signals the determination on the part of policy makers to find the ‘right’ conditions that will make this digestible to the public. Neoliberal politics and New Public Management ideology of public administration have captured the imagination of policy makers in matters of technology innovation and management. They have made for social, political and financial pressures on state institutions to try to generate value out of any viable public asset. But the protracted occupation by private sector of public infrastructure provision has the effect of stifling imagination and discussion of alternative ways to develop public technology (Sharon 2020; Tempini 2022). For a long time, scholarship has further pointed out how technology infrastructure and standards have an inertia that makes radical revision as difficult as function creep is easy (Hanseth and Lyytinen 2010). Once important social and organisational interdependencies have developed over ICT that serves as infrastructure, it is very difficult to unseat it. From the care.data debacle (REF) to the latest GDPR move, NHS data have been one of the fields in which the pressure to data function creep is highest, raising the appeal for government to try to avoid the turbulence of inevitable questioning by trying to get plans pass scrutiny as quickly and silently as possible. As the Ada Lovelace Institute observes, this fits a pattern of “*decide, announce and defend*” approach to the public focused on its persuasion instead of its involvement (Machirori and Patel 2021; Ada Lovelace Institute 2020). It is a self-defeating strategy, as each time, the plans have been halted by public debate that is only in part focused on the merits and demerits of the envisioned ways of using data and doing research with them; and is focused, in the rest, on public issues of governance, accountability, and trust. The GDPR’s roll out has now been indefinitely frozen (O’Donovan 2021), as the size of deliberate opt-outs from a distrusting public was concerning. The pandemic has offered a few tell-tales on the issues and risks of data function creep. The UK government attempted to extend regulations of pandemic emergency that

loosened constraints on the access to sensitive health data (Tempini 2022; O'Donovan 2021), but its failure to attend to public trust led to new failure (Wilson 2021; Bharti et al. 2021).

The public deserves much better, more responsible approaches to data governance accountability and control of function creep. The public should be involved in the deliberation and articulation of values and aims embedded in projects and infrastructures using data about it (Bharti et al. 2021). Projects should be paramountly concerned with gaining public trust. At a minimum, Kitchin recommends that: *Citizens should know precisely what the app seeks to achieve and what will happen with their data. There should also be safeguards to stop control creep and the technology being repurposed for general or national security, predictive policing or other governance or commercial purposes*” (Kitchin 2020). And project managers should be concerned about the cumulative effects that straining public trust with exploitative or unaccountable data management can have in the long-term. In pandemic times, such risks are accelerated as function creep can ensue very quickly. An example is the Covid Symptom Study, data self-reporting app ZOE, launched by a private company and academics at the start of the first wave. Its funding was quickly secured, for the ensuing 2 years, by the government. Only few months later, when the research team started publishing promising results over the power of self-report symptom data to predict infections, authors observed the trend for the ZOE data to be *“increasingly being linked to the public health response within the National Health Service (NHS) in the United Kingdom”* (Drew et al. 2020). In the immediate, this could be seen as a good example of collective mobilisation and ingenuity in the face of exceptional challenge. But it poses concerns over the risks this shift could mean for the livelihood of individuals (See Box 1 ‘Zoe’).

These are not only ethical concerns. They often become epistemological if people start to take countermeasures which in turn affecting the quality and reliability of the data (Bharti et al. 2021). As about any health researcher knows, opt-outs are not randomly distributed. Just as many other individual preferences and social facts, they introduce bias in experimental and observational research (Teira 2013). When sizeable, they quickly lead to issues statistical representation of the remaining sample. Thus, Bharti and colleagues (2021) stress how public involvement in the articulation of values and aims of a data project is not only politically beneficial. It leads to better outcomes. The GDPR is a case in point, as the value of the proposition was tarnished by huge spikes in drop-out rates (O'Donovan 2021). And as the controversies around the ‘pingdemic’ demonstrated, allowing an app to report information

can give a 3rd party the power to interfere with our everyday life at a latter point, on terms one might not be happy with. Perhaps ironically, many deleted the contact-tracing app once it started doing what it was supposed to do and ‘pinged’ them out of circulation. Contact-tracing apps are designed to discipline and reshape spatial movements and social connections (Kitchin 2020), but they have an uneven impact. Essential workers are likely to face higher costs of ‘protecting others’ policies, if the government does not provide for any other forms of support. The complex dynamics of social systems (Wilson 2021) mean that these workers will have very high incentives to flout unfair rules. Kitchin also notes that data function creep can also be indirect. Location data that is automatically generated, for public health apps, by a smartphone operating system’s location services, can be shared by the operating system with other apps that also use location data; with the result of deepening the surveillance of a user within the ecosystem of data brokers and surveillance capitalism.

2.2 Consent and privacy

Research with Internet data has been raising ethical concerns related to consent and the distribution of harm and benefits for quite some time. Data science and AI methods often rely on the use of large amounts of real-world data to train machines, and these data have often been taken where they are most easily available, e.g. much of the Internet. Internet users directly generate data as they post, record, interact over platforms, but data are also continuously generated by the underpinning technological infrastructures as a matter of mere technical operation. The whole array of data generation and storage instances is impossible to keep track of (Clark et al. 2019). Intense debates have been generated over the ethical standards that new forms of digital research should adhere to (Petermann et al. 2022); including questions over the double standards that public- vs private-sponsored research follow. Questions of consent will be important for self-report data collection initiatives. If data could be reused multiple times a whole host of questions are raised as to what could be the appropriate mechanisms for granting consent for uses other than those originally envisaged at the point of data generation.

Calls for ‘blanket’ consent that pre-emptively authorise a broad spectrum of re-use (and the milder version of ‘broad’ consent) have been heavily criticised for their vulnerability to misuse, capture and manipulation. Also heavily undermined have been individualistic arguments that envision each research participant deciding on the use of their own data as the most desirable ideal of data governance – most impractical but also unethical to dump all the

responsibility onto isolated individuals. Organisational arrangements that require more complex governance procedures and accountability structures have been better received.

Self-report data, especially when health-related, are often very sensitive, giving rise to the potential for direct and indirect harm from misuse, manipulation and profiling. It is thus important for contributors to be able to keep track and access contextually relevant information as to what is done with their data, and what is envisioned for the future. They should also be made aware that parties interested in accessing and using their data, because ethical standards and guidelines as to what are admissible re-uses of individual data change over time and across research domains and institutions, might not always think that contributors have a right to informed consent when:

- the data are shared in a public setting seemingly without whatsoever expectation of privacy (for instance, tweets from public accounts)
- there are no expectations that participants could be directly harmed from the research (for instance through anonymisation of the data at the point of collection)
- when they are working with a private sector company that has served legally-compliant notifications – for instance, a fair processing notice that is GDPR compliant can pre-empt, at once, a whole host of uses and projects the company will undertake with the data.

Given the higher exposure to diverse and distributed expectations and engagements, high standards should be required of initiatives relying on self-report data. To be meaningful, question of consent should be renewed at major project milestones or updates. Participants should be actively invited to consider whether changing situations fit them, and should not be expected to keep track and make sense of change by themselves. For instance, if the governance of a project changes due to change in ownership or management, self-report data should not be considered a conventional asset that can be sold on. Instead, they should be treated as objects that bind different people together in an economy of relational ethics (Prainsack 2019a; Prainsack 2019b; Birhane 2021). Data generation and access should be minimised as much as possible. Many questions require functioning ethical oversight.

2.3 Contribution and participation

Self-report data collection initiatives can achieve impressive feats thanks to the economics of crowdsourcing, where a large number of contributors each invest only a small portion of their time to a great cumulative effect. However, it is not easy to run these initiatives over a long

time, and how to do this will be particularly crucial for those initiatives whose data collection becomes more valuable the more it has taken place regularly and over a long period of time, e.g. data about health events and symptoms used for the understanding the spread of disease.

The emergency of the pandemic, with the emotional response and mobilisation it generated, along with the lockdowns and the way many people redirected their attention and free time to activities that can be completed at home, created one of the easiest scenarios for a self-report data collection project to succeed. Many people wanted to understand and learn about c19, and needed reassurance and explanations on their daily experience. Many wanted to help. It was easier to accept intrusions of privacy the boundaries and consequences of which were not fully clear. And still, even in the times of pandemic, several challenges arose to the motivation and distribution of participation. The UK ‘pingdemic’, and the consequent migration of many users deleting the contact tracing app from their phone, spoke of an unresolved tension on expectations and boundaries of the data reporting that contributors are voluntarily enabling; a tension that was only exacerbated by the experimental nature of the methods involved.

A rich literature on participation in data self-reporting offers many points of concern to keep in mind for future self-reporting projects. The seminal ‘ladder of participation’ published by Sherry R. Arnstein (Arnstein 1969) has since served to highlight how nuanced the concept of participation can be, open to manipulation and as a result, how appearances can be misleading. An extensive literature has followed to question the concept of participation further. Self-reporting projects distributed over the web are often opaque as to their inner, complex workings, while very public about the moral economy that they want to draw on to mobilise support, with frequent calls to share for the common good, for altruism, emancipation, empowerment of oneself and their kin. On close scrutiny, participatory practices can turn out to be empty, tokenistic, or extractive ‘crowdsourcing’. Kelty and colleagues (Kelty et al. 2015) point out that participation on the web would be better understood as something that is not a linear spectrum along one axis, but rather as something multi-dimensional. In this way, the limited openness to participation of most web projects is easier to observe, with most projects restricting participation to one or few dimensions.

It is crucial that self-report data collection projects respect the effort and investment put in by contributors to make it possible by relating to them as invaluable partners and project

stakeholders. Extractive models popular in the Silicon Valley, that promote the value of sharing and promise empowerment and emancipation, only to extract information from individuals while excluding them from benefits (H. R. Ekbia 2016) and governance of a project, will come under fire and will be increasingly unpalatable, given the recent mood change in public sphere discussions on these matters. Participation in health research involves a particular kind of labour that involves turning one's body into readable and available to observation and cognition (Brives 2013; Milne 2018; Cooper 2012); and even in the case of unambiguously commercial projects where crowds are called on to contribute cognitive labour in exchange for financial retribution, as in Amazon's Mechanical Turk, strong arguments have been raised that point to the exploitation that these models build on (H. Ekbia and Nardi 2014; Irani and Silberman 2013; Nardi 2015).

It is beyond dispute that many projects have exploited contributors who are asked to volunteer time and information while they are kept out of any relevant sense of ownership of the results. In light of much literature it is clear that it is difficult to develop a self-report data collection project without controversy. The asymmetries of projects with a very large, distributed base of contributors and a very small team of managers and developers will make for very sensitive politics of contribution and participation. Managers of data self-reporting projects should avoid exploiting a common double standard where the efforts of the contributor base are celebrated and recognised with the language of empowerment and 'bottom-up', or 'patient-led', research while at the same time, individual contributors are excluded from formal recognition when the research is published in peer-review outlets. Only a few projects have tried to recognise the individual contributors that a crowd is made of through authorship credits, and this not always be possible nor desirable. There are many other ways to formally recognise the contribution of the public in these projects other than scientific credit. Direct involvement in governance and management can be an alternative, as well as having formal and public ways to gather and respond to request, motions and value demands.

It is important, and even more so once out of pandemic time, for self-report data collection projects to consider very seriously how they can best and actively engage in transparent communication about all aspects of the project, including decision structures and finalities (See Box 2 'COVID-HEAL'). The projects that do not cause controversy are likely to be those that engage in a sustained dialogue with the base of contributors; explain key changes

in governance and direction of the project; demonstrate accountability to the contributor base at least as keenly as they strive towards other stakeholders such as sponsors, 3rd party researchers or regulators.

2.4 Self-report and self-harm

Data can be used for purposes other than initially envisaged in many different ways. Contributors themselves, once engaged and active in the data collection, can repurpose or imitate the methods and inclusive rhetoric of the projects they joined, and extend aims of the data collection and networking taking place on a self-report data collection system to organise research on matters of their own interest.

There is a long history of this kind of initiatives in health research, with perhaps the most historically significant one that captured by the pioneering work by Steven Epstein (Epstein 1996), describing how in the midst of the US AIDS epidemic, ACT-UP activists aiming to influence research on HIV and extend access to experimental therapies harnessed existing networks of civil rights activists to coordinate independent research, protest and sabotage.

Since the rise of the web, techniques and methods for organising independent research that exploits available resources, data and technology have grown more sophisticated. Examples have multiplied as we saw the emergence of self-report data collection beyond the domain of scientific research and into the domains of civic and political action, where data have been used by activists to mobilise evidence in support of their claims (Milan and Velden 2016; Bruno, Didier, and Vitale 2014); as well as artistic performances and other individual applications of data collection on the conduct of one's life, in the culture of the Quantified Self movement (Sharon 2017). In the field of health research, the self-organisation of data collection through web networks and technology has been called 'patient-led research' (Vayena and Tasioulas 2013) in the wake of some celebrated examples (Paul Wicks et al. 2008; Paul Wicks et al. 2011) that excited commentators about the social web's potential in scientific production.

Exciting as it might seem for ordinary individuals to take knowledge matters into their own hands, there are a number of thorny ethical issues associated with distributed contributors' ability to organise data collection on the side (Vayena et al. 2015; Tempini and Teira 2019; P. Wicks, Vaughan, and Heywood 2014; Ledford 2018). It is clear that people enrolled on these

initiatives often interpret the data thus generated in order to make life decisions with potentially large consequences. For instance, patients might make changes in their treatment course, or take uncontrolled chemicals. And when contributors are many, different aims, expectations, levels of engagements and degrees of literacy on the subject matter are at play. Some will be much more vulnerable than others. They have many questions that they might seek to satisfy with what they find, whether the answer is provided through state of the art methods and project level supervision or not.

These projects completely lack of ethical supervision, let alone of the sort that has long been required of publicly funded university research. No one will have a view as to who these people might be and how they should be supported. What these initiatives risk creating then is something short of well-developed solidarity, but rather, an ephemeral alliance and confluence of interests between individuals who might then be find themselves to face the consequences alone. The regulatory trend towards allowing patients more freedom to experiment with treatment courses means that the risk calculus is becoming more complex just as patients gain more freedom to do it themselves (Carrieri, Peccatori, and Boniolo 2018; Navarro, Tempini, and Teira 2021; Tempini and Teira 2020).

Pandemic time is only likely to exacerbate the problem. With c19 we have seen the emergence in social media and public sphere of a number of theories, since debunked, as to the causes of illness and more or less implausible treatments that are within easy reach of each citizen with a curious mind and misplaced skepticism. Many people are willing to shape actions and medical decisions based on information they find more or less casually. They are mostly on their own evaluating risks and kinds of harms that could arise. Spontaneous and ephemeral instances of self-report data collection are likely to compound this kind of issue, by buying credibility, and time, to implausible theories and methodologies. In several examples, patients eventually self-harm (Paul Wicks, Heywood, and Vaughan 2012).

Self-harm is not the only relevant ethical issue. Another important issue is the potential for contributor-led initiatives to interfere with other, better designed clinical research, such as clinical trials, when participants decide to tamper with the experimental protocol by carrying out their actions on the side from it. The result could be the creation of noise, the jeopardisation or slowing down of the best research. These issues can create a rather uncomfortable situation for the managers and developers of the technology platforms these

initiatives are relying on. An issue for them will also be how liabilities and best courses of action can be worryingly unclear.

2.5 Making the data worth it

Given that contributors can invest a significant amount of resources labouring to collect data (Milne 2018; Cooper 2012), develop expectations, and can face risk of undesirable consequences of participating in the data collection, it is really important that data are valued and used. As it has been made obvious by great amounts of literature, much of the value that has been ascribed to new kinds of data-intensive data collection, including the collection of self-report data, is due to the assumed possibility of reusing the same data multiple times for different purposes. Given that data can acquire new value when they are linked or juxtaposed with other data and questions, their value could be renewed as many times as people having access to them believe it is worth to do so.

While the investment on the part of an individual contributor can be variable and small, the cumulative investment asked of society, or a relevant group within it, can be considerable; the more so the more contributors a project is able to attract. A self-report data collection project, therefore, will face demands of accountability as to the ways in which the data have been put to use; and whether the amount of effort it commanded was worth the good it generates.

If little is done with datasets that people might have a reasonable expectation more could have been done with, a tension can arise. Scholarship on data self-reporting has thus questioned (Prainsack 2017; Tempini and Del Savio 2019; Sharon 2016) specific projects that harness the rhetoric of empowerment and participation but might not live up to expectations (see Box 3 ‘PLM’); and asked if those who directly contributing are then excluded from data governance decisions of the data, the incentive to maximise the use of the data might be weaker (Birhane 2021; Prainsack 2019b; Hafen 2019; Tempini and Del Savio 2019). Expert practitioners involved in health data self-reporting studies often stress the duty that a project has to make justice to the participant’s burden.

There are various reasons why a dataset could not be widely re-used. For instance, in the case of health data, their sensitive nature and the high risk of misinterpretation require not everybody should be given access. Also, the data that is actually collected might not be as good and reliable as initially expected by the researchers who design the exercise. Much

depends on the everyday practical circumstances of a self-selected group of contributors, with different expectations, hopes and levels of literacy.

Health data can be particularly expensive to keep and govern. They require complex infrastructures to protect the data from unauthorised access. This means keeping up with continuously changing security standards and a quickly evolving risk and threat landscape; while at the same time, maintaining the knowledge that is required to ensure that the data remain valuable, of high quality, and that their peculiarities and qualities are well understood by those who directly re-use them (Demir and Murtagh 2013; Tempini and Leonelli 2018). In this respect, self-report data pose more challenges because they are often made available in a structural lack contextual of information and awareness of the specific situation in which they were generated by unknown, distributed volunteer contributors. Also, many data self-reporting projects, especially when funded through public research funds, can run into issues of long-term maintenance and sustainability; it is important that plans are made early on as to how the data will be managed in the long run and how continuous funding and support for data governance functions should be secured.

At any rate, data self-reporting projects should attend to questions of contributor benefit; reflecting on how to feed back relevant and valuable information learned from the research done with the data, a task that becomes more difficult the wider and more diverse the contributor base is; actively rejecting forces pulling the relationship with contributors towards an extractive state, where contributors keep being asked to contribute data while a clear sense of benefit and utilisation is being lost. This attitude requires active vigilance on the part of project managers and a willingness to challenge abnormal power asymmetries, especially when the project is driven by private entities. Powerful tech giants have entrenched their economic might, to the detriment of the public good, on extractive relationships where ever more data is continuously collected in order to predict and generate behaviour.

Acknowledgements

I would like to acknowledge the input of Paul Wicks, PPI for COVID-HEAL, who offered time to chat on the topic as an expert practitioner; and thank Cian O'Donovan, James Wilson and Melanie Smallman for feedback and suggestions.

BOXES

BOX 1: ZOE

The ZOE COVID Study is a self-report study of covid symptoms and patient experience centred around a smartphone app developed by a company, ZOE, and a team of nutrition and epidemiology researchers at King's College, London. The app itself is an example of the speed at which existing technology, code and methods can be redeployed for new purposes. As the first wave hit the UK, the company refactored the app it had been developing for the study of nutrition into an app that could be used by covid patients to record their symptoms on a daily basis, for months, so as to be able to estimate the dynamics of the disease as it swept through society. ZOE and the KCL team led by Tim Spector are serial data self-report study leaders, having been known as The British Gut project, where they asked participants to provide funding, samples and data in exchange for personalised information and advice on gut health. The participation of contributors was already limited to the completion of a few well circumscribed tasks, with no direct input in project management and day to day operations; but it was still taken to be one of a crop of web-based projects that were breaking boundaries towards science democratisation (Del Savio, Prainsack, and Buyx 2016).

When ZOE COVID launched it was quickly downloaded by a great amount of people, thanks to the extraordinary momentum granted by the pandemic emergency (Varsavsky et al. 2021; Drew et al. 2020). It was also adopted in other countries (Kennedy et al. 2021). Contributors input demographic data upon registration, then are asked to submit a daily report of symptoms and covid test results. The data could then be used by researchers wanting to study covid patient experience; and estimate its demographic and geographic distribution as well as geographic and case count movements. Studies reported the self-reported data could be used to produce estimates that closely track those gained from best methods, leading researchers to conclude that data self-report can be a valuable complementary resource of data to be used by policy makers and public health officers in coordinating emergency response. They should be particularly useful, they add, to understand the situation in regions where best method estimates are not viable due to lack in testing capacity.

The success of the ZOE Covid app in providing a resource of good quality data was outstanding, but there remain open questions as to how feasible (and desirable) an initiative of such proportions would be outside pandemic time. The initiative commanded such extraordinary participation thanks to the extraordinary context in which many people found

themselves; and benefited from the lockdowns keeping the population at home, and constant media coverage that kept the issue relevant and into focus for the many users that the app needs to keep daily reporting even when they feel well (to provide a baseline of negative symptomatology). It also benefited from sudden financial backing from the government, as the app was, in the researcher's own words, "*increasingly being linked to the public health response within the NHS*" (Drew et al. 2020). As researchers observed (Varsavsky et al. 2021)(p5), relatively high numbers of weekly active users are needed to confidently detect relatively small rises in case counts.

However, there are various limitations and experts have been split over how much confidence can be put in these data to guide emergency response. Study researchers observe how the self-selected sample of respondents provides for a biased and non-representative sample of the population, with a trend towards more representation from the less vulnerable. It fits the experience of self-reporting initiatives in general. Events like the 'pingdemic' that unexpectedly requested so many citizens to self-isolate after technology had determined they had been exposed to a positive carrier could also suddenly pull the rug from under such initiatives. But other interventions can also interfere, such as vaccinations, or being clear of covid after infection, can provide a sense of security that makes some less interested. The same ZOE technology applied to Sweden, a country that had famously resisted the introduction of social distancing measures and lockdowns, had seen a lower uptake and again biased in favour of better-served areas (Kennedy et al. 2021) clustered around the universities running the study. It is potentially a warning sign for an approach that is recommended as valuable complement in the management of the emergency in the less well resourced region of a country. People who have worse access to testing might be disadvantaged twice, if the surrogate measures also risk being worse.

Worries of function creep and surveillance would not be easily assuaged by sudden partnerships with state institutions (Drew et al. 2020). Emergency response is a highly dynamic situation and even when the consequences of measurement are not taken directly to specific individuals, when public health officials are trying to coordinate a response which might include various measures affecting people's livelihoods, we are already in a situation where self-report data have been repurposed to a new use and become what they were not initially meant to, e.g. a tool for population management. Experts warned about the potential emergence of incentives to lie; of worries of being reported to the police and in general of

risk contributing data to the app; the potential consequences of these worries for the quality of data collection and of other related activities such as self-administered tests; and the consequences of the latter on the quality of policy making.

BOX 2: COVID-HEAL

HEAL-COVID is a study of post-hospitalisation covid symptoms and long-term outcomes. At the time of writing, it reports enrolment of over 1000 patients who had been hospitalised with covid. Patients are asked to self-report data through a questionnaire that has been designed over only a few weeks by an interdisciplinary team of specialists, including specialists of patient participation. The urgency of the pandemic with its hospitalisation surges threatening NHS standards of care had a capacity for accelerating collaboration and overcoming challenges such as questionnaire copyright licensing, that in normal times can hold up the development of an experiment for much longer. Some licenses were obtained at a very quick turnaround. To find at record speed the patient representatives that could offer review and support in research design, the team was able to make use of Cambridge University hospital's group. Patient involvement in questionnaire design focused the team's efforts on designing a minimalist questionnaire of a minimum number of items – it was important not only to ask patients for feedback over questionnaire and workflow drafts, but also to have an open space where the patient could ask the questions they had come up with.

Experts point out to contributor burden as both an ethical and epistemological challenge. Wasting time and energy of patients who are already challenged by a debilitating disease, many of whom have long-covid, should be avoided. But also, it is an epistemological challenge because over-burdening contributors diminishes the quality of responses, especially so when many patients are suffering from fatigue over a long time. This meant various trade-offs in deciding what symptoms patients should be asked about. What is important to patients from an experiential point of view might be different from other symptoms who could be warning signs of dangerous complications. Tests that are very valuable but very difficult to self-report well (e.g. a 'six minute walk test') could be weighed against other less informative symptoms that are more intuitive and familiar to patients, such as mood changes. From a self-reporting perspective, reporting on symptoms that are important to patients and that patients can at the same time report on reliably was a key principle of questionnaire drafting.

As they designed the study at the beginning of the pandemic, researchers had to make assumptions about the kind of patient that would be most likely to participate in the experiment. Experts note how the composition of the team, and the ecology of actors around them (funders, academics) was not representative of society's demography. All things being equal, this can make the research more vulnerable to biases and assumptions baked in the research design and questionnaires. They also learned that the infection moved through society in unpredictable and ever changing ways. While the first wave saw ethnic minorities and essential workers such as taxi drivers disproportionately hospitalised, by the time the third wave was sending patients into hospitals, they were disproportionately the unvaccinated. The consequences of these shifts for the validity of the questionnaire and how its items had been selected were unclear.

It also makes more difficult to elaborate and take actions from the feedback about the study that is gathered from the participants; this will also come with hidden biases and assumptions, as researchers collect information only from the self-selected few who accepted to participate. Expert found it much more challenging to know about those who are not participating in the study – more likely to be vaccine and covid sceptics. It is unclear what are the consequences of collecting worse data about specific social groups. It is clear that much depends on how consequential the actions that can be taken from the data that are collected can be – a reason for caution when imagining the use of self-reported data in policy-making and crisis response.

BOX 3: PLM

PatientsLikeMe (PLM) is a social media network and platform of online communities centred on the patient experience. It gathers hundreds of thousands of patients suffering from thousands of different conditions. The patients gather to socialise, learn from one another and about their condition, and participate in health research. The platform, unlike most social media, is free of ads. The for-profit company sells research services and access to pseudonymised patient data. Its researchers have published many peer reviewed publications, some very celebrated for the way they leveraged the web to produce scientific knowledge faster and cheaper than traditional methods. They made PLM one of the most hyped and promising social media in the health space. PLM promised a revolution of health research and care, a model of patient empowerment that would allow them to rewrite the research agenda of pharmaceutical sector, democratising health. This was hoped for especially in

respect to ‘orphan diseases’ – conditions that, usually because of low patient numbers, fail to attract the level of investment and research necessary to develop treatments that make an impact. PLM was started as a community for ALS patients by a family affected by the condition who had been at the centre of some of the most dynamic activism and research around the disease – called ‘guerrilla scientists’ their early efforts had already attracted, before the founding of PLM, the attention of a Pulitzer-winning journalist (Weiner 2004); and documentarists who took their story to Sundance (Ascher and Jordan 2006).

The reality of everyday operations at PLM was more challenging than many commentators excited about what the social web could mean for knowledge production might have assumed. The collection of self-reported data from such a disparate base of contributors posed threw up many data quality issues related to the effort to bridge between the world of patient experience, knowledge, language, aims and expectations, and the world of standardised scientific observations, recording and communication, of data structures and taxonomies (Arnott-Smith and Wicks 2008; Tempini 2015; Paul Wicks et al. 2010; Frost and Massagli 2008).

The need to collect data that would be worth, at once, for patients and their personal sense of biographical trajectory; and for third parties interested in learning about them for socialisation, scientific or business purposes led to conflicting demands. At play were different definitions of what is a valuable direction for platform and what is necessary burden and attrition (Tempini 2017). The platform was committed both to a bottom-up revolution of the health research industry; and to the development of a viable, self-sustaining business model in a highly competitive and dynamic health industry, where executive board members and venture capitalists have a way to focus one’s mind. Its monopolistic control over the self-report data, with little direct participation by patients in day-to-day operations and decisions, created an unresolvable tension (Tempini and Del Savio 2019), and ambiguity towards some genuinely spontaneous patient-led initiatives that threatened to perturb the overall design (Tempini and Teira 2019).

References

- Ada Lovelace Institute. 2020. *The Data Will See You Now: Datafication and the Boundaries of Health*. 978-1-8382567-0-8. London: Ada Lovelace Institute.
- Ada Lovelace Institute and The Health Foundation. 2021. *The Data Divide: Public Attitudes to Tackling Social and Health Inequalities in the COVID-19 Pandemic and Beyond*. 978-1-8382567-3-9. London: Ada Lovelace Institute.
- Arnott-Smith, Catherine, and Paul Wicks. 2008. 'PatientsLikeMe: Consumer Health Vocabulary as a Folksonomy'. *AMIA Annual Symposium Proceedings 2008*: 682–686. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2656083/>.
- Arnstein, Sherry R. 1969. 'A Ladder Of Citizen Participation'. *Journal of the American Institute of Planners* 35 (4). Routledge: 216–224. doi:10.1080/01944366908977225.
- Ascher, Steven, and Jeanne Jordan. 2006. *So Much, So Fast*. DVD. West City Films.
- Bharti, Nishtha, Cian O'Donovan, Melanie Smallman, and James Wilson. 2021. 'Public Trust, Deliberative Engagement and Health Data Projects: Beyond Legal Provisions'. *Engaging Science, Technology, and Society* 7 (1): 125–133. doi:10.17351/ests2021.1197.
- Birhane, Abeba. 2021. 'Algorithmic Injustice: A Relational Ethics Approach'. *Patterns* 2 (2): 100205. doi:10.1016/j.patter.2021.100205.
- Brives, Charlotte. 2013. 'Identifying Ontologies in a Clinical Trial'. *Social Studies of Science* 43 (3): 397–416. doi:10.1177/0306312712472406.
- Bruno, Isabelle, Emmanuel Didier, and Tommaso Vitale. 2014. 'Statactivism: Forms of Action between Disclosure and Affirmation'. *Partecipazione e Conflitto* 7 (2): 198–220.
- Carrieri, D., F. A. Peccatori, and G. Boniolo. 2018. 'The Ethical Plausibility of the "Right To Try" Laws'. *Critical Reviews in Oncology/Hematology* 122 (February): 64–71. doi:10.1016/j.critrevonc.2017.12.014.
- Clark, Karin, Matt Duckham, Marilys Guillemain, Assunta Hunter, Jodie McVernon, Christine O'Keefe, Cathy Pitkin, et al. 2019. 'Advancing the Ethical Use of Digital Data in Human Research: Challenges and Strategies to Promote Ethical Practice'. *Ethics and Information Technology* 21 (1): 59–73. doi:10.1007/s10676-018-9490-4.
- Cooper, M. 2012. 'The Pharmacology of Distributed Experiment - User-Generated Drug Innovation'. *Body & Society* 18 (3–4): 18–43. doi:10.1177/1357034X12446380.
- Del Savio, Lorenzo, Barbara Prainsack, and A. M. Buyx. 2016. 'Crowdsourcing the Human Gut. Is Crowdsourcing Also "Citizen Science"?' *Journal of Science Communication* 15 (03): A03 1-16. http://jcom.sissa.it/archive/15/03/JCOM_1503_2016_A03?utm_content=buffer34a84&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer.
- Demir, Ipek, and Madeleine J. Murtagh. 2013. 'Data Sharing across Biobanks: Epistemic Values, Data Mutability and Data Incommensurability'. *New Genetics and Society* 32 (4): 350–365. doi:10.1080/14636778.2013.846582.
- Drew, David A, Long H Nguyen, Claire J Steves, Cristina Menni, Maxim Freydin, Thomas Varsavsky, Carole H Sudre, et al. 2020. 'Rapid Implementation of Mobile Technology for Real-Time Epidemiology of COVID-19', 7.
- Ekbja, Hamid, and Bonnie Nardi. 2014. 'Heteromation and Its (Dis)Contents: The Invisible Division of Labor between Humans and Machines'. *First Monday* 19 (6). <http://journals.uic.edu/ojs/index.php/fm/article/view/5331>.

- Ekbia, Hamid R. 2016. 'Digital Inclusion and Social Exclusion: The Political Economy of Value in a Networked World'. *The Information Society* 32 (3): 165–175. doi:10.1080/01972243.2016.1153009.
- Epstein, Steven. 1996. *Impure Science: AIDS, Activism, and the Politics of Knowledge*. London: University of California Press.
- Financial Times*. 2022. 'Palantir Gears up to Expand Its Reach into UK's NHS', June 9.
- Frost, Jeana H, and Michael P Massagli. 2008. 'Social Uses of Personal Health Information Within PatientsLikeMe, an Online Patient Community: What Can Happen When Patients Have Access to One Another's Data'. *Journal of Medical Internet Research* 10 (3). doi:10.2196/jmir.1053.
- Hafen, Ernst. 2019. 'Personal Data Cooperatives – A New Data Governance Framework for Data Donations and Precision Health'. In *The Ethics of Medical Data Donation*, edited by Jenny Krutzinna and Luciano Floridi, 141–149. Philosophical Studies Series. Cham: Springer International Publishing. doi:10.1007/978-3-030-04363-6_9.
- Hanseth, Ole, and Kalle Lyytinen. 2010. 'Design Theory for Dynamic Complexity in Information Infrastructures: The Case of Building Internet'. *Journal of Information Technology* 25 (1): 1–19. doi:10.1057/jit.2009.19.
- Irani, Lilly C., and M. Six Silberman. 2013. 'Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 611–620. CHI '13. New York, NY, USA: Association for Computing Machinery. doi:10.1145/2470654.2470742.
- Kelty, Christopher, Aaron Panofsky, Morgan Currie, Roderic Crooks, Seth Erickson, Patricia Garcia, Michael Wartenbe, and Stacy Wood. 2015. 'Seven Dimensions of Contemporary Participation Disentangled'. *Journal of the Association for Information Science and Technology* 66 (3): 474–488. doi:10.1002/asi.23202.
- Kennedy, Beatrice, Hugo Fitipaldi, Ulf Hammar, Marlena Maziarz, Neli Tsereteli, Nikolay Oskolkov, Georgios Varotsis, et al. 2021. 'App-Based COVID-19 Surveillance and Prediction: The COVID Symptom Study Sweden'. medRxiv. doi:10.1101/2021.06.16.21258691.
- Kitchin, Rob. 2020. 'Civil Liberties or Public Health, or Civil Liberties and Public Health? Using Surveillance Technologies to Tackle the Spread of COVID-19'. *Space and Polity* 24 (3). Routledge: 362–381. doi:10.1080/13562576.2020.1770587.
- Lanzing, Marjolein. 2021. 'Contact Tracing Apps: An Ethical Roadmap'. *Ethics and Information Technology* 23 (1): 87–90. doi:10.1007/s10676-020-09548-w.
- Ledford, Heidi. 2018. 'How Facebook and Twitter Could Be the next Disruptive Force in Clinical Trials'. *Nature* 563 (7731): 312. doi:10.1038/d41586-018-07351-8.
- Lucivero, Federica, Nina Hallowell, Stephanie Johnson, Barbara Prainsack, Gabrielle Samuel, and Tamar Sharon. 2020. 'COVID-19 and Contact Tracing Apps: Ethical Challenges for a Social Experiment on a Global Scale'. *Journal of Bioethical Inquiry* 17 (4): 835–839. doi:10.1007/s11673-020-10016-9.
- Lucivero, Federica, Luca Marelli, Nora Hangel, Bettina Maria Zimmermann, Barbara Prainsack, Ilaria Galasso, Ruth Horn, et al. 2021. 'Normative Positions towards COVID-19 Contact-Tracing Apps: Findings from a Large-Scale Qualitative Study in Nine European Countries'. *Critical Public Health* 0 (0). Taylor & Francis: 1–14. doi:10.1080/09581596.2021.1925634.
- Machirori, Mavis, and Reema Patel. 2021. 'Turning Distrust in Data Sharing into "Engage, Deliberate, Decide"'. Blog. *The Ada Lovelace Institute Blog*.

- <https://www.adalovelaceinstitute.org/blog/distrust-data-sharing-engage-deliberate-decide/>.
- Milan, Stefania, and Lonneke van der Velden. 2016. 'The Alternative Epistemologies of Data Activism'. *Digital Culture & Society* 2 (2): 57–74. doi:10.14361/dcs-2016-0205.
- Milne, Richard. 2018. 'From People with Dementia to People with Data: Participation and Value in Alzheimer's Disease Research'. *BioSocieties*, January, 1–17. doi:10.1057/s41292-017-0112-x.
- Morozov, Evgeny. 2014. *To Save Everything, Click Here: The Folly of Technological Solutionism*. PublicAffairs.
- Nardi, Bonnie. 2015. 'Inequality and Limits'. In . Irvine, CA. http://wtf.tw/ref/nardi_inequality.pdf.
- Navarro, David Fraile, Niccolò Tempini, and David Teira. 2021. 'The Trade-off between Impartiality and Freedom in the 21st Century Cures Act'. *Philosophy of Medicine* 2 (1). doi:10.5195/philm.2021.24.
- O'Donovan, Cian. 2021. 'Debate Needed on Post-Pandemic Rules for Medical Data'. *Research Professional News*. <https://www.researchprofessionalnews.com/rr-news-political-science-blog-2021-12-debate-needed-on-post-pandemic-rules-for-medical-data/>.
- Petermann, Mylene, Andrew Strait, Niccolò Tempini, Ismael Kherroubi-Garcia, and Kirstie Whitaker. 2022. *Ethical Review Processes for AI and Data Science Research*. London and Exeter: Ada Lovelace Institute, University of Exeter Institute for Data Science and Artificial Intelligence, Alan Turing Institute.
- Prainsack, Barbara. 2017. *Personalized Medicine: Empowered Patients in the 21st Century?* New York: New York University Press.
- Prainsack, Barbara. 2019a. 'Data Donation: How to Resist the Leviathan'. In *The Ethics of Medical Data Donation*, edited by Jenny Krutzinna and Luciano Floridi, 9–22. Philosophical Studies Series. Cham: Springer International Publishing. doi:10.1007/978-3-030-04363-6_2.
- Prainsack, Barbara. 2019b. 'Logged out: Ownership, Exclusion and Public Value in the Digital Data and Information Commons'. *Big Data & Society* 6 (1): 2053951719829773. doi:10.1177/2053951719829773.
- Rabeharisoa, Vololona, Tiago Moreira, and Madeleine Akrich. 2014. 'Evidence-Based Activism: Patients', Users' and Activists' Groups in Knowledge Society'. *BioSocieties* 9 (2): 111–128. doi:10.1057/biosoc.2014.2.
- Sharon, Tamar. 2016. 'The Googlization of Health Research: From Disruptive Innovation to Disruptive Ethics'. *Personalized Medicine*, October. doi:10.2217/pme-2016-0057.
- Sharon, Tamar. 2017. 'Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare'. *Philosophy & Technology* 30 (1): 93–121. doi:10.1007/s13347-016-0215-5.
- Sharon, Tamar. 2020. 'When Google and Apple Get Privacy Right, Is There Still Something Wrong?' *Medium*. <https://medium.com/@TamarSharon/when-google-and-apple-get-privacy-right-is-there-still-something-wrong-a7be4166c295>.
- Shirky, Clay. 2010. *Cognitive Surplus*. London: Penguin Press. <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/1846142180>.

- Teira, David. 2013. 'Blinding and the Non-Interference Assumption in Medical and Social Trials'. *Philosophy of the Social Sciences* 43 (3): 358–372. doi:10.1177/0048393113488871.
- Tempini, Niccolò. 2015. 'Governing PatientsLikeMe: Information Production and Research through an Open, Distributed and Data-Based Social Media Network'. *The Information Society* 31 (2): 193–211. doi:10.1080/01972243.2015.998108.
- Tempini, Niccolò. 2017. 'Till Data Do Us Part: Understanding Data-Based Value Creation in Data-Intensive Infrastructures'. *Information and Organization* 27 (4): 191–210. doi:10.1016/j.infoandorg.2017.08.001.
- Tempini, Niccolò. 2022. 'Pandemic Data Circulation (Crossing Boundaries)'. *Tecnoscienza* 13 (1). doi:10.5281/zenodo.6627015.
- Tempini, Niccolò, and Lorenzo Del Savio. 2019. 'Digital Orphans: Data Closure and Openness in Patient-Powered Networks'. *BioSocieties* 14 (2): 205–227. doi:10.1057/s41292-018-0125-0.
- Tempini, Niccolò, and Sabina Leonelli. 2018. 'Concealment and Discovery: The Role of Information Security in Biomedical Data Re-Use'. *Social Studies of Science* 48 (5): 663–690. doi:10.1177/0306312718804875.
- Tempini, Niccolò, and David Teira. 2019. 'Is the Genie out of the Bottle? Digital Platforms and the Future of Clinical Trials'. *Economy and Society* 48 (1): 77–106. doi:10.1080/03085147.2018.1547496.
- Tempini, Niccolò, and David Teira. 2020. 'The Babel of Drugs: On the Consequences of Evidential Pluralism in Pharmaceutical Regulation and Regulatory Data Journeys'. In *Data Journeys in the Sciences*, edited by Sabina Leonelli and Niccolò Tempini, 207–225. Cham: Springer International Publishing. doi:10.1007/978-3-030-37177-7_11.
- Varsavsky, Thomas, Mark S. Graham, Liane S. Canas, Sajaysurya Ganesh, Joan Capdevila Pujol, Carole H. Sudre, Benjamin Murray, et al. 2021. 'Detecting COVID-19 Infection Hotspots in England Using Large-Scale Self-Reported Data from a Mobile Application: A Prospective, Observational Study'. *The Lancet Public Health* 6 (1). Elsevier: e21–e29. doi:10.1016/S2468-2667(20)30269-3.
- Vayena, Effy, Roger Brownsword, Sarah Jane Edwards, Bastian Greshake, Jeffrey P. Kahn, Navjyot Ladher, Jonathan Montgomery, et al. 2015. 'Research Led by Participants: A New Social Contract for a New Kind of Research'. *Journal of Medical Ethics*, March, medethics-2015-102663. doi:10.1136/medethics-2015-102663.
- Vayena, Effy, and John Tasioulas. 2013. 'The Ethics of Participant-Led Biomedical Research'. *Nature Biotechnology* 31 (9): 786–787. doi:10.1038/nbt.2692.
- Vezyridis, Paraskevas, and Stephen Timmons. 2017. 'Understanding the Care.Data Conundrum: New Information Flows for Economic Growth'. *Big Data & Society* 4 (1). SAGE Publications Ltd: 2053951716688490. doi:10.1177/2053951716688490.
- Weiner, Jonathan. 2004. *His Brother's Keeper: A Story from the Edge of Medicine*. New York, NY: Harper Collins. <http://litmed.med.nyu.edu/Annotation?action=view&annid=12306>.
- Wicks, P., T. Vaughan, and J. Heywood. 2014. 'Subjects No More: What Happens When Trial Participants Realize They Hold the Power?' *BMJ* 348 (jan28 9): g368–g368. doi:10.1136/bmj.g368.
- Wicks, Paul, James Heywood, and Timothy Vaughan. 2012. 'Waiting for P<0.05'. figshare. http://figshare.com/articles/Waiting_for_p<0.05/96802.

- Wicks, Paul, M. Massagli, J. Frost, H. Macedo, K. Felzer, and J. Heywood. 2008. 'A Patient-Led Trial of Lithium in ALS Using the Internet. *Amyotroph Lateral Scler* 2008;9:S59'. *Amyotrophic Lateral Sclerosis* 9 (sup1): 59–66. doi:10.1080/17482960802444469.
- Wicks, Paul, Michael P Massagli, Jeana Frost, Catherine Brownstein, Sally Okun, Timothy E Vaughan, Richard Bradley, and James Heywood. 2010. 'Sharing Health Data for Better Outcomes on PatientsLikeMe'. *Journal of Medical Internet Research* 12 (2): e19. doi:http://dx.doi.org/10.2196%2Fjmir.1549.
- Wicks, Paul, Timothy E Vaughan, Michael P Massagli, and James Heywood. 2011. 'Accelerated Clinical Discovery Using Self-Reported Patient Data Collected Online and a Patient-Matching Algorithm'. *Nature Biotechnology* 29 (5): 411–414. doi:10.1038/nbt.1837.
- Wilson, James. 2021. 'The Intervening State'. *The RSA*.
<https://www.thersa.org/comment/2021/09/the-intervening-state>.
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.