

Regulation and use of confidential patient information for genomic and medical research during and post COVID-19

FUNDED BY

NIHR | National Institute
for Health Research

Authors

Colin Mitchell, Tanya Brigden, Sarah Cook and Alison Hall

November 2021

A PHG Foundation report on the regulation and use of confidential patient information for genomic and medical research during and post COVID-19

Acknowledgements

This report is independent research commissioned and funded by the Department of Health and Social Care Policy Research Programme (*How effective are changes to information governance improving access to confidential patient data and/or personal data for medical research purposes? Should these changes be made permanent in the future?* NIHR202463). The views expressed in this publication are those of the authors and not necessarily those of the Department of Health and Social Care.

We are very grateful to the members of our External Advisory Panel, Professor Anneke Lucassen, Dr Christine Patch, Rebecca Asher and Dr Jeffrey Skopek for their guidance and feedback on this research.

We thank the interviewees named in Appendix 1 for generously providing their time and expert insights on this topic. We are also indebted to our collaborators, Dr Zoe Khor and Hannah Kearsey, from Traverse Limited who organised and ran our focus group, and to Dr Malcolm Oswald, Director of Citizens Juries c.i.c., who provided helpful insights for this element of the research.

Disclaimer

The following report is intended to provide general information and understanding of the law. It should not be considered legal advice, nor used as a substitute for seeking qualified legal advice.

URLs in this report were correct as of October 2021

Written and produced by PHG Foundation

2 Worts Causeway, Cambridge, CB1 8RN, UK +44 (0)1223 761900

www.phgfoundation.org

© 2021 PHG Foundation

Correspondence to:

intelligence@phgfoundation.org

The PHG Foundation is a health policy think-tank and linked exempt charity of the University of Cambridge. We work to achieve better health through the responsible and evidence-based application of biomedical science. We are a registered company, no. 5823194.

Executive summary

At the start of the COVID-19 pandemic, in March 2020, the government in England introduced measures to enable the use of confidential patient information (CPI) for COVID-19 purposes without consent or another form of approval that would normally be required. These measures, the 'COPI notices', set aside the common law duty of confidentiality for a range of purposes, including research into the disease and its impact on health and care. This report considers how these regulatory changes to the governance of confidential patient information have impacted genomic and medical research, and whether these changes should be integrated into the regulatory framework longer-term.

The significant impact of the COPI notices on genomic and medical research

To assess the impact of the COPI notices we reviewed the landscape of data use and linkage for genomic and medical research during the COVID-19 pandemic. We also interviewed key stakeholders from research, public health and population-level data initiatives. We identified a considerable range of data initiatives addressing COVID-19 that have relied (at least in part) on the COPI notices. Many have leveraged existing projects, infrastructures and organisations in order to address COVID-19. They include the COVID-19 Genomics UK (COG-UK) consortium, initiatives established by Public Health England for genomic surveillance, and large-scale genomic research initiatives, such as the GenOMICC study - with linkages to Genomics England and COG-UK, the HOCl study initiated by COG-UK, the SIREN study and research enabled by UK Biobank.

The evidence suggests that the COPI notices have had significant positive impacts in terms of the speed and efficiency of data access for research and also in improving access to sources of data, such as primary care data, which had been hard to obtain prior to the pandemic. The notices are likely to have had an impact in several ways: introducing a new regulatory pathway for research without consent or approval from the NHS Health Research Authority - to enable or mandate disclosure of CPI for COVID-19 purposes; development of new or streamlined processes to facilitate COPI notice-authorized data access and; the powerful signal they have sent about the importance of data sharing and access to combat COVID-19. Untangling these elements will be important in determining whether, and what manner of, changes should be taken forwards on a permanent basis.

Extension of the COPI notices or continued exceptions for COVID-19

COVID-19 still presents a major threat. Ongoing surveillance and research will be necessary to manage infection levels and assess the risk of new mutations or variants for some time. While this is necessary, it could be argued that measures to facilitate processing of confidential patient information for COVID-19 purposes should remain in place. However, the impact of vaccination may have altered the equation and it

is clear from our focus group and from wider empirical research of public attitudes in England, that the public desire greater transparency about what is being done with CPI and justifications for any changes. The NHS Health Research Authority has published guidance for the transition of research reliant on the COPI notices to the conventional pathway for approval of research using CPI without consent based on a recommendation from the Confidentiality Advisory Group. This might signal a return to normal for most research but these measures could remain in place for specific forms of processing or flows of data between specific actors. If longer term exceptions are made, we recommend that greater transparency is provided about the scope of 'COVID-19 purposes' and the oversight of an independent body is considered.

Building on the COPI notices with further reforms

Our work has highlighted the need to build on the experience of facilitating fast and efficient access to confidential patient information for research in the public interest, while maintaining high levels of public and professional confidence. The COPI notices may already have catalysed improvements through streamlining of processes and the strong signal of support for health data sharing that they engendered, but calls for further regulatory changes could be envisaged. For example, the mandatory sharing of certain categories of data, such as primary care data, or mandatory sharing with specific recipients such as Trusted Research Environments, for research purposes. Further work and consultation is required to determine whether such reforms are necessary and proportionate. Our reviews of the legal framework and health data landscape, interviews with key stakeholders and research on public attitudes to data sharing do however, highlight a set of key ethical and legal considerations that should be taken into account in the development of any proposals for reform.

Ethical and legal considerations for changes to regulation of CPI for health research

The central pillar for the use of CPI for research and other secondary purposes is the trust and confidence of professionals, patients and publics in the institutions, processes and individuals involved. Our research highlights widespread agreement with Onora O'Neill's account, that those responsible must endeavour to demonstrate that they are worthy of patients' and publics' trust. A number of factors will be important in demonstrating such **trustworthiness**.

Transparency is crucial and a desire for greater transparency has been repeatedly emphasised in empirical work through the pandemic. This does not simply mean provision of information about the potential benefits of data sharing but there should be clarity about the scope of potential uses, the nature of the data involved, safeguards, residual risks and justifications, including the opportunity cost of not sharing data. However, although necessary, provision of information is not in itself sufficient. There is also an imperative for public **involvement and engagement** in decisions about CPI.

Our research identifies a desire from both the public and professionals that greater efforts are made to engage with communities and groups in both general and specific decisions about the use of CPI. The

General Practice Data for Research and Planning initiative has provided a timely example that inadequate transparency and engagement can set-back plans to unlock the power of data for health. Millions of patients have subsequently exercised their National Data Opt-out (NDO) and opted-out entirely from the use of data for secondary purposes.

Consent and choice have an important role in the health data system. The NDO is a relatively blunt instrument because it does not allow individuals to set their preferences about specific forms of research or recipients they are comfortable with, the opt-out is all or nothing. However, as a policy it was felt to strike the right balance in supporting autonomy without placing a burden of myriad specific decisions on individuals or the development of complex systems to facilitate and maintain downstream commitments on the health service and researchers. The system is not perfect and there are challenges which would benefit from further consideration. In particular, developing consensus between healthcare professionals who owe a duty of confidentiality to their patients and researchers obtaining specific informed consent for use of their confidential patient information. This is likely to require agreement on the information required to demonstrate and communicate the scope and validity of such consents to data custodians.

It is no surprise that concerns of **privacy, data security and data protection** are foremost among the risks perceived by the public in the use of patient data for research. The development of Trusted Research Environments that enable highly secure and de-identified processing of data for research without sharing data is highly promising in this regard. Empirical work during the pandemic found high levels of approval for data processing via OpenSAFELY, a platform developed by the DataLab at the University of Oxford for secure analysis of patient records.

A perennial challenge highlighted by our legal analysis and interviews is the complexity of the regulatory landscape. The interaction between different regulatory domains can be confusing and uncertain for professionals as well as the public. In particular, **data protection law and the common law duty of confidentiality** overlap significantly but they have important differences in terms of scope and legal requirements. Further guidance for professionals on this topic would be useful and there should be efforts to explain aspects, such as the nature of 'pseudonymised' data and how it differs from 'anonymised' data to patients and the public.

A second area of overlap is between **health research** and **public health surveillance**. Under the COPI notices certain activities may not have had to categorically differentiate between these fields and it may be appropriate that the requirements for each should be aligned where they overlap. However, it will be important to ensure that regulatory reforms do not inadvertently exclude certain actors or inappropriately incentivise the use of one route over another.

The social licence in a time of change

Finally, these considerations and decisions about reform are part of a broader dynamic context. Rapidly advancing technologies, including artificial intelligence, new models of healthcare, such as the learning healthcare system, the integration of diverse sources of data—from genomic and ‘omic’ data to data generated through wearable devices and apps—hold great promise and also significant challenges for regulation, governance and the relationship between the patient and the healthcare system. It may be time for a broad public dialogue, with engagement at local and national levels, about uses of health data and the social contract or licence that underpins it. New models of governance, such as data trusts or intermediaries, and continued technical developments may have an important role to play. Ultimately, unlocking the power of data for public benefit will only be achieved through commitments to transparency and consultation with the public about how data are used.

Table of contents

Executive summary.....	2
1. Introduction.....	8
1.1 Description of the scientific, regulatory and policy environment pre COVID-19.....	8
1.2 Scope and methodology.....	10
1.3 Section summary	13
2. The legal framework governing the use of confidential patient information for medical research.....	14
2.1 Overview.....	14
2.2 The common law duty of confidentiality (CLDC).....	16
2.3 The Health Service (Control of Patient Information) Regulations.....	19
2.4 Data protection law	25
2.5 Distinguishing between surveillance and research	27
2.6 The National Data Opt-out	28
2.7 Section summary	30
3. The landscape of data use for genomic and medical research during COVID-19	32
3.1 Introduction.....	32
3.2 The health data landscape	32
3.3 Genomic surveillance	39
3.4 Genomic research reliant on COPI notices.....	43
3.5 The future of COVID-19 research and surveillance	45
3.6 Section summary	47
4. Patient and public attitudes to the use of confidential data for research	48
4.1. Introduction.....	48
4.2. Focus group	48
4.3. Key considerations affecting public attitudes towards data sharing in a pandemic.....	50
4.4. The future of the COPI notices	60
4.5. Section summary	61
5. Discussion	62

5.1	How have regulatory changes to information governance to support research into COVID-19 impacted genomic and medical research?	62
5.2	Should part, or all, of the changes be permanently integrated into the UK's regulatory framework?	64
6.	Conclusions.....	73
	References	76
	Appendix 1 – Interviewees	88
	Appendix 2 – Exemplar research studies	89
	Appendix 3 – Traverse focus group report	95

1. Introduction

1.1 Description of the scientific, regulatory and policy environment pre COVID-19

This report evaluates the impact of emergency regulatory measures introduced to facilitate the use of confidential patient information in response to the COVID-19 pandemic. These measures (the COPI notices) were introduced pursuant to the Control of Patient Information Regulations (2002). This set of Regulations was introduced in 2002 to facilitate sharing confidential patient information when justified for certain purposes including public health and cancer treatment, and include a power for the Secretary of State to issue a notice, to mandate sharing when other situations arise. Notices under these Regulations were issued in March 2021 and justified on the basis that they were needed to manage the UK's responses to a novel threat – the SARS-CoV-2 virus. Since March 2020 there has been an unprecedented focus on how research (characterised as 'science') could and should inform the public health and governmental response. This report focuses on a small but essential part of that research – genetic and genomic research – and the extent to which the COPI notices facilitate such research, the regulatory and ethical issues arising as a consequence of this decision, and provides some considerations and recommendations for the future.

Before analysing the COPI notices and their impact in more detail, it is helpful to reflect on the role of confidential patient information in health more generally. Confidential patient information underpins the health and social care system. The confidential information shared by patients with their health and social care professionals informs both individual patient care and management, but also the planning, management and functioning of health systems and services. Yet the governance of that confidential patient information is often inconsistent and fragmented despite efforts at standardisation. Unlike personal data, which is regulated by statute through the UK General Data Protection Regulation 2016 and the UK Data Protection Act 2018, confidential patient information is primarily governed through the common law informed by case law through the courts, and influenced by professional practice.¹

Recognition of the fundamental importance of confidential information in health care led to publication of a set of overarching principles in 1997 to provide additional guidance to practitioners about developing best practice.² These principles were promulgated by the National Data Guardian for Health and Social Care, Dame Fiona Caldicott, who initially had a non-statutory role in providing advice to the health and social care sector about their use of confidential patient information. These principles were updated in 2013 pursuant to an Information Governance Review.³ Publication of Data Security Standards⁴ reinforced the importance of technical and procedural standards in keeping health information safe and secure. The singular role of confidential data in health was given statutory weight through the Health and Social Care (National Data Guardian) Act 2018 which entrenched the role and influence of the National Data Guardian as a key source of authoritative guidance.

Despite these efforts to create harmonised approaches, the governance of confidential patient information remains highly fragmented and takes place at a local level where the duty of confidentiality resides between individual patients and professionals. The Caldicott Guardian system is a key part of the decision making framework. This requires all NHS organisations and local authorities which provide social services to have a senior person responsible for protecting the confidentiality of people's health and care information and to make sure that it is used properly. The UK Caldicott Guardian Council (a subgroup of the National Data Guardian's Panel)⁵ provides an informal source of advice and guidance for these professionals, but is not a professional body. The decisions made by these professionals are influenced by their professional background and experience, and unlike other professional roles, accountability for decision making as part of this role is sometimes unclear given the lack of a regulatory body for Caldicott Guardians.

The initial focus of the Caldicott Principles was on highlighting that confidential patient information should be handled securely and responsibly, and the potential harms associated with inadvertent or reckless disclosure. However, it became clear that the failure to disclose confidential patient information when the use was justified and proportionate, could also result in significant harm. The Control of Patient Information Regulations (2002) were enacted, in part, to provide a statutory route for health professionals to disclose certain types of confidential information, such as information about suspected or confirmed cancer diagnoses, or information about infectious diseases, without fear of recrimination or even litigation. These regulations constituted statutory recognition that, in some cases, the public interest in sharing specific confidential data outweighs the public interest in keeping those data confidential. Subsequent efforts of the National Data Guardian to highlight the importance of disclosing confidential patient data, by adding an additional Caldicott Principle highlighting the importance of a 'duty to share information for direct care'⁶ only appeared to have limited impact on changing professional practice.

Commentators,⁷ and successive government departments⁸ have highlighted that unlocking patient records through greater digitisation and adoption of harmonised standards for recording clinical symptoms in a systematic manner⁹ could yield substantial benefits for publics and for the country more generally. These include creating a richer dataset which could be utilised for more effective planning and management, but also potentially as a resource for public and private research.

Despite these objectives, efforts to streamline the collection and use of confidential patient information have faced considerable challenges. Some of these have been infrastructural, such as the poor technical infrastructure supporting digital services within and between NHS organisations. Others have been more philosophical, questioning the nature and form of the social contract between citizens, the NHS and government, and the reciprocal rights and obligations arising over that confidential information. Some efforts--such as the care.data initiative--to streamline sharing of confidential patient data have encountered a negative response from some publics and professional groups.¹⁰ Other initiatives have been

criticised for a lack of clarity about what is at stake, the rights and duties that are engaged, aggravated by opaque terminology.¹¹

It comes as no surprise that prior to the COVID-19 pandemic, researchers found this complex, multi-layered regulatory landscape difficult to navigate. Streamlining the Integrated Research Application System (IRAS) to bring together parallel applications for ethical approval (covering data protection, clinical trial and medical device aspects) with approval from the Confidentiality Advisory Group for the use of confidential patient data without consent went some way to simplify the application process. But researchers continued to experience difficulties in accessing data, particularly from primary care providers. They were also frustrated by long delays, inconsistent decision making from ethics committees and other gatekeepers. A report exploring researchers' experiences chronicled difficulties with consent and confidentiality, arising from 'disproportionate information governance' and 'unfounded concerns' held by the data providers about the legitimacy of processing. These difficulties were compounded by researchers' reports that the recipients of requests for data appeared to lack the requisite expertise to apply data protection requirements in a proportionate and workable way.¹²

The emergence of the SARS-CoV-2 virus in early 2020 represented an unprecedented global threat to public health. In response, the introduction of the COPI notices in March 2020 to suspend or expedite the approval mechanisms to the use of confidential patient data, was couched as a proportionate reaction to this global emergency.

1.2 Scope and methodology

Scope

Set against this backdrop the aim of our research was to answer two related questions:

1. How have regulatory changes to information governance to support research into COVID-19 impacted genomic and medical research?
2. Should part, or all, of the changes be permanently integrated into the regulatory framework?

The relatively limited debate concerning the use of confidential data from individuals which was generated when these notices were introduced was in stark contrast with the vigorous debate about the potential negative impact of the General Data Protection Regulation on genomic and medical research prior to its implementation in 2018. As part of this work, we wanted to explore the role of public trust and trustworthiness, to understand whether the notices were regarded as a proportionate response to an overwhelming threat, or whether the lack of response might signify a more permanent change in public opinion.

All confidential patient information is potentially highly sensitive but we have a particular focus on uses and linkages of genomic data during the pandemic as it is a form of data which is viewed as both highly disclosive and highly identifying. Our scientific review highlights the considerable activity in this area during the pandemic but genomic data is rarely used in isolation from other clinical data and our legal analysis, consideration of patient/public attitudes, interviews and ethical analysis all have implications for processing confidential patient information for research purposes more generally.

This report focuses on the impact of the law governing confidential patient information in England. We have not addressed the devolved nations specifically, since health is a devolved matter and separate arrangements apply to Scotland, Northern Ireland and Wales. For this reason, our scientific review, focus group and interviews are all limited to the position in England

Methodology

Scientific review of exemplar COVID-19 scientific and clinical research

In order to address the first question, we analysed a number of metrics as a proxy for the nature and volume of the research that was facilitated as a result of the COPI notices. These included reviewing the measures that were put in place to systematically collect viral and human genetic/genomic data (section 3.2); investigating how these data were released/shared with researchers and the conditions of their release (section 3.3); and analysing exemplar scientific and clinical research on SARS-CoV-2 and COVID-19 to explore these in more detail including evaluating public facing statements about reliance on COPI notices as a legal basis for disclosure (section 3.4). Exemplar projects were selected by reviewing scientific and genomic literature to assess how patient data, human and pathogen genetic/genomic data had been accessed, processed and integrated to generate research results. In each case we examined study methodology, study outputs (including blogs and commentaries in the public domain), and in some cases interviews with researchers and policy makers (see Appendix 1).

Taken together, these measures allowed us to assess the nature and the volume of the research that was done and make inferences about how the notices had facilitated the use of confidential patient information.

An analysis of the information governance changes (Legal/regulatory analysis)

The significance and effect of the COPI notices and supporting measures were evaluated through desk-based analysis. For legal sources we have utilised legal search engines including Lexis Library and Westlaw UK, and University of Cambridge resources to access relevant primary and secondary legislation.

Ethical and societal analysis

Relevant ethical and societal literatures were analysed through desk-based review, again utilising resources via the University of Cambridge. This work was also informed by the findings of two five-day citizens' juries which were held in Spring 2021 as this project got underway. Led by Dr Sabine van de Veer, these focused on health data sharing in a pandemic. They considered the future of candidate pandemic data initiatives which influence direct care, service planning and research, such as the NHSX COVID-19 Data Store.¹³ Dr Oswald provided early sight of the outputs from these citizens' juries which assisted in our research on ethical and societal factors and in planning for the public/patient focus group.

Independent patient/public focus group

In order to inform our understanding of patient and public views on the impact of the COPI notices, we commissioned an independent collaborator - Traverse Ltd - to organise, convene, facilitate and report on a dedicated patient/public focus group to explore public views around the processing of confidential data without consent (section 4.2; Appendix 3). The two hour focus group consisted of ten volunteers, who had at least one interaction with the NHS over the preceding year. Researchers from PHG Foundation worked with Traverse to design the focus group content, materials and semi-structured discussion questions. The group was facilitated independently and results were fed back by the end of week 8 to inform subsequent stages of the project (namely the legal and ethical analysis, interviews and input from the Expert Advisory Panel).

Interviews

In addition to the desk-based analysis, we conducted 13 interviews to supplement and enrich our research findings. Our interviewees were purposively selected to ensure representation from key stakeholder groups, and also, in some cases, to address particular points of uncertainty or contention. Each interview was conducted using a semi-structured list of questions tailored to the interviewee's interest and expertise. At least two researchers from PHG Foundation were involved in each interview, which were recorded for note-taking purposes. Key themes/findings were extracted and are reflected in this report, but specific findings are not attributed to specific interviewees as these were carried out under Chatham House Rules. A list of interviewees is included in Appendix 1.

Expert Advisory Panel

An external Expert Advisory Panel was convened to advise on the project scope, delivery and outputs. This comprised of independent experts who have relevant knowledge and/or policy experience including a representative from Genomics England Ltd with expertise on genomic research and data governance, a legal academic from the Centre for Law Medicine and Life Sciences with expertise on medical innovation

and privacy, an expert on patient and participant views of data use and governance and a clinical representative from the Nuffield Council on Bioethics 'Biological and Health Data' working group.¹⁴ The group met three times during the seven month long project, at weeks 4, 20 and 24. At the first meeting, the group deliberated the research design and planned delivery; at the second meeting, they considered the findings of the independent focus group and their impact on the research, and at the final meeting they reviewed the draft report and had an opportunity to highlight topics for inclusion in the discussion and recommendation sections of the report.

1.3 Section summary

This section explores the regulatory landscape prior to the COVID-19 pandemic, highlighting some of the challenges associated with the use of confidential patient information. It then describes the scope of this project, the research questions that have been addressed and the methodology used to generate the findings in this report.

In the following section, we provide a detailed evaluation of the legal framework governing the use of confidential patient information for medical research invoked as an emergency response to the SARS-2-CoV virus, namely the Control of Patient Information Notices published in the Spring of 2020.

2. The legal framework governing the use of confidential patient information for medical research

In this section, we outline the legal framework governing the use of confidential patient information for genomic and medical research. Our focus is on research use but we touch on relevant legal considerations governing the use of such information for public health purposes (in so far as they may overlap with research use). We describe the key regulatory changes adopted as emergency measures to streamline data processing, sharing and linking for COVID-19 purposes – the ‘COPI notices’ – and we situate them in the wider legal framework governing health data in the UK. This provides a grounding for the following sections which describe the landscape of key data flows and initiatives for genomic and medical research during the pandemic and report the findings of our own focus group, alongside other research, on public attitudes to the use of patient data for research during the pandemic and into the future. In section 5 we combine these elements to discuss key legal and ethical issues, picking up on some of the legal challenges outlined in this section.

2.1 Overview

Unlike some other nations, the United Kingdom has no specific legislation directly governing genomic research or genomic data. Instead the protection and disclosure of genomic data, patient data and other relevant data for genomic and medical research is governed by a range of legislation and court-based law (common law). This involves formal regulation by the courts and regulators, such as the Information Commissioner’s Office (ICO), as well as less formal regulation and oversight from a range of advisory bodies, health or research authorities and professional bodies, such as the National Data Guardian for Health and Social Care (NDG), and the NHS Health Research Authority (HRA). There are differences in the legal framework across the four nations of the UK. In this research we are focused primarily on the law of England and Wales and the regulatory and governance framework in England.

Table 1 sets out key parts of this framework and our research is focused primarily on changes in relation to the law of confidentiality and the disclosure of confidential patient information during the COVID-19 pandemic.

Table 1: Legal areas impacting use of patient information for research purposes

Legal area	Relevance
Right to privacy under Article 8 of the European Convention on Human Rights	Incorporated into UK law by the Human Rights Act 1998. This provides a right to privacy in relation to personal information for all citizens and the courts must consider this right when determining any case relating to private information, even if framed in one of the legal areas below.
Tort of misuse of private information	Recognised as a distinct tort or legal wrong in 2014. This tort is similar to but distinct from the common law of confidentiality with different tests and subject matter.
Common law duty of confidentiality	An 'equitable' cause of action that has historically focused on unauthorised use of information which has a 'quality of confidence' and is imparted in circumstances importing an obligation of confidence. As discussed in this section, this has evolved considerably in light of the Human Rights Act.
Data protection law	The UK GDPR and Data Protection Act 2018 set a range of rights and obligations in relation to 'personal data' (which do not need to be private in nature). The requirements of data protection law do not always neatly align with the other areas of law above.

Our review of these legal areas has identified that the key regulatory change to information governance introduced in the pandemic has been the notices made pursuant to the Control of Patient Information Regulations 2002 for COVID-19 purposes (the COPI notices). We have not identified other relevant regulatory changes, although guidance has been issued in relation to data protection law in particular without a change in the underlying law.

Therefore, our focus is on the application of the common law duty of confidentiality to genomic and medical research. However, as outlined above, this is only one part of the relevant legal framework so these changes cannot be evaluated completely in isolation. We will discuss aspects of the legal areas set out in the above table where they have an important interaction with confidentiality in the medical research context. First we begin with a description of the common law duty of confidentiality (CLDC) and its application in the context of genomic and medical research.

2.2 The common law duty of confidentiality (CLDC)

Confidentiality is a cornerstone of healthcare and biomedical research¹⁵ and it is central to maintaining the trust and confidence of patients and participants in healthcare and research. The essence of the ethical and legal duty of confidentiality is that confidential information should not be disclosed without authorisation from the individual or in another legally recognised form. In terms of the legal duty, the courts have allowed claims for breach of confidence for over 150 years. Over this time, a three part test for a claim of breach of confidence has been established. This asks:

- Does the information have the necessary quality of confidence?
- Was the information imparted in circumstances importing an obligation of confidence?
- Has there been an unauthorised use of the information causing detriment?¹⁶

The implementation of the Human Rights Act in 1998 led to the law of confidentiality being interpreted and applied in light of the right to respect for private and family life enshrined in Article 8 of the European Convention on Human Rights, and the jurisprudence developed by the European Court of Human Rights.¹⁷ This means that, in the case of personal information (as opposed to confidentiality arising from trade secrets for example), the courts will first ask whether the circumstances give rise to a reasonable expectation that privacy will be protected.¹⁸ This is extremely likely to be the case in the health and medical context. The courts have emphasised that the details of one's medical circumstances are 'obviously private'¹⁹ and deserve the full protection afforded by the law of confidence (subject to the interpretation given to it by the Human Rights Act 1998)—even if medical details are revealed in a public place. Because it is so well established that the doctor-patient relationship gives rise to an obligation of confidence, the key question for disclosure of confidential genomic and health information for research is how disclosure may be authorised.

2.2.1 Lawful disclosure of confidential information

To the extent that the information is private and confidential, any disclosure must be lawful and in accordance with the patient's right to privacy and confidentiality. The NHS has a Confidentiality Code of Practice (although it has not been updated since 2003) which sets out different approaches to the use of confidential information depending on the purpose.²⁰ This approach is also followed by the GMC's 2017 guidance, Confidentiality: good practice in handling patient information.²¹ These draw on the approach of the courts to set out three key ways in which disclosure of confidential patient information may be lawful:

- Explicit or implied consent to disclosure;
- Disclosure in the public interest;
- Authorisation in law.

Consent to disclosure of confidential information

The simplest way of authorising disclosure of confidential information is through explicit consent. This requires explanation of what will be done with the information and an agreement either orally or in writing. However, it is not entirely clear what nature and level of information is required. There has been very little legal discussion of the informational requirements of a valid consent for the disclosure of confidential information—as opposed to consent under data protection law or informed consent to medical treatment.²² At a minimum it is clear that there must be adequate information that explains in broad terms how confidential information will be used. However, as Chico and Taylor suggest,²³ this is not the same as importing the more extensive requirements for ‘informed consent’ to medical treatment established in the law of medical negligence, for example.

An alternative option is implied consent to sharing confidential information. The GMC guidance states that ‘implied consent refers to circumstances in which it would be reasonable to infer that the patient agrees to the use of the information, even though this has not been directly expressed’.²⁴ As Taylor and Wilson discuss, implied consent is (or should be) a ‘real’ consent that respects individual autonomy.²⁵ Compared with ‘explicit consent’ it is only the signal of consent that is different. It should be based on the same standards of information and voluntariness. An implied consent is different to hypothetical or presumed consent because it is signalled through conduct.

The implication of the GMC guidance and NHS Code of Practice is that, in some circumstances, it may be acceptable to assume that consent is implied by the patient’s continued acceptance of care.²⁶ What has been most challenging is determining in which circumstances and relation to what purposes, it may be acceptable to assume that consent is implied by a patient. The GMCs guidance is that this may be the case in relation to ‘direct care’ or ‘clinical audit’²⁷ but making these designations is not always straightforward. At present, implied consent is not relied on for disclosure of confidential patient data in the NHS for wider purposes.

Finally, the GMC requires that a patient has ‘ready access’ to information about what will be done with such information, for example in leaflets, posters, on websites and face-to-face. As well as this, the GMC guidance is that for a valid consent to be implied, the patient must have an opportunity to object. Such an objection should be respected unless there is an overriding ‘public interest’ (see below) or the patient lacks capacity and the decision is in line with their overall best interests.

A reasonable expectation approach

Mark Taylor and James Wilson have proposed an alternative way of recognising when confidential medical information may be disclosed; by determining when disclosure is within the 'reasonable expectations' of the patient.²⁸ This concept is not alien to the courts because they have already established that, in assessing whether a right to privacy is engaged, the key question is whether there is a reasonable expectation of privacy in the circumstances.²⁹ Taylor and Wilson suggest that it is logical that there are circumstances where there has been no invasion of privacy because the disclosure of information is within the 'reasonable expectations' of the patient. They argue that this could be the case where there are sufficient indications of a respect for individual autonomy and dignity (including signs of notification and acceptance) and that the courts have adopted an objective approach to assessing reasonable expectations of privacy. This means that neither the recipient of the confidential information, nor the patient's *actual* expectations will determine the matter. Instead, the courts adopt an 'objective' perspective and will consider the issue from the perspective of a hypothetical 'reasonable person' who is in the position of the patient. From this perspective, the courts will take a wide range of factors and considerations into account when they determine whether an expectation of privacy is *reasonable* in the circumstances. Taylor and Wilson argue that this should include the significance or sensitivity of the information, whether there has been notification and acceptance by the patient and any research that provides a more in depth understanding of patients' expectations.

As Taylor and Wilson propose, this could be approached like the concept of 'no surprises' using two questions:

- What has been done to ensure that people have reason to expect this use?
- What has been done 'to ensure that they accept it'?³⁰

The status of this approach to confidentiality is currently unclear but it is a potentially important reframing which was explored by the previous National Data Guardian, Dame Fiona Caldicott.³¹

Disclosures in the public interest

Confidentiality is not absolute, so without express or implied consent it may be possible to lawfully disclose information as long as it is justified in the public interest for important public benefits. Similarly, the right to private life in Art 8 European Convention on Human Rights is not absolute, and is qualified by interference which is necessary in a democratic society to protect objectives such as health or the rights and freedoms of others. Any public interest justification has to meet a very high threshold of not only outweighing a patient's interest in autonomy and confidentiality but also the public interest in maintaining confidentiality

of medical information. The GMC guidance makes clear this will only apply in exceptional circumstances such as when there is a significant risk of serious harm.³²

Statutory basis

Finally, and most importantly for this research, confidential information may be lawfully disclosed if there is a statutory basis (sometimes described as a 'gateway') setting aside the duty of confidentiality for the disclosure. Although this is set out as a separate route to lawful disclosure, these statutory gateways may also be described as setting out circumstances in which it would be lawful to disclose confidential information in the public interest. The most important of these for our research are found in section 251 of the National Health Service Act 2006, and the Health Service (Control of Patient Information) Regulations 2002. These 'COPI' regulations specify a process for approval of research using confidential patient data without consent, which pre-dates the COVID-19 pandemic, and also provide a basis for the COPI notices introduced during the COVID-19 pandemic.

2.3 The Health Service (Control of Patient Information) Regulations

In this section we consider the potential statutory gateways for processing confidential patient information for genomic and medical research purposes and for public health or pandemic response purposes. The Health Service (Control of Patient Information) Regulations 2002 provide key routes for the disclosure of confidential patient information for research purposes.³³ The basis for these routes to disclosure is found in section 251 of the National Health Service Act 2006, which provided the Secretary of State ("SoS" - for Health and Social Care) with the power to regulate (and mandate where necessary) the processing of 'patient information' for 'medical purposes', if they consider it is necessary or expedient in the interests of improving patient care, or, in the public interest. 'Medical purposes' are defined very broadly and encompass, diagnosis, medical research, planning and preventative medicine.³⁴ Patient information is also defined very broadly to include 'information (however recorded) which is to any extent derived, directly or indirectly, from information which relates to the physical or mental health or condition of an individual, to the diagnosis of his condition or to his care or treatment'.³⁵ Critically, this includes the ability to make regulations and require the processing of 'confidential patient information' (known as 'CPI').

Confidential Patient Information (CPI)

Our research is focused on the use of confidential patient information (CPI) for genomic and medical research. This is because this is the category of data which has been impacted by the COPI notices designed to help respond to the COVID-19 pandemic.

CPI is defined by s251(11):

‘For the purposes of this section, patient information is “confidential patient information” where—

- (a) the identity of the individual in question is ascertainable—
 - (i) from that information, or
 - (ii) from that information and other information which is in the possession of, or is likely to come into the possession of, the person processing that information, and
- (b) that information was obtained or generated by a person who, in the circumstances, owed an obligation of confidence to that individual.’

The definition is important because it sets the scope for the ensuing rules and regulations that apply to research and other forms of processing of confidential patient information. Our analysis will return to this point but for now it is useful to keep in mind that this definition may not *automatically* correspond to the scope of confidential patient information as interpreted by the courts nor does it match the scope of ‘personal data’ under the UK GDPR.

Powers to process CPI under the COPI Regulations

Within these parameters a range of rules and provisions for the processing of patient information are set out in the Health Service (Control of Patient Information) Regulations 2002, which extend to England and Wales. Different parts of these Regulations provide for different forms of processing, some of which are particularly relevant to this research.

Crucially, Regulation 4 of the Control of Patient Data Regulations sets aside the common law duty of confidence for any processing under the regulations: ‘Anything done by a person that is necessary for the purpose of processing confidential patient information in accordance with these Regulations shall be taken to be lawfully done despite any obligation of confidence owed by that person in respect of it.’³⁶

2.3.1 Research under Regulation 5

For researchers, the avenue for lawful research using CPI without consent prior to the COVID-19 pandemic was via Regulation 5. This enables processing for ‘medical purposes’ (including medical research) in the circumstances set out in the schedule to the Regulations, but only provided they are approved by the NHS Health Research Authority and a research ethics committee in the case of medical research or, in any other case, by the Secretary of State.

The Health Research Authority was obliged³⁷ to appoint a committee to provide advice in relation to applications for approval under this regulation, known as the Confidentiality Advisory Group or CAG. This group provides advice and detailed recommendations to the HRA for REC approved research applications

and also scrutinises applications for non-research purposes to make recommendations to the Secretary of State. Applicants are advised to use the HRA's online tool³⁸ to help determine whether their proposal is research. Because research can only be approved within the scope of s.251, researchers must demonstrate that:

- it is not possible to carry out the activity another way, taking into account cost and available technologies (s.251(4)). This means that the CAG will have to be satisfied, based on the circumstances, that seeking consent is neither possible nor practical.
- there must be a clear indication that the proposal is in the public interest or will improve patient care. This means that a public benefit will be required for research applications. (s.251(1)(a-b))
- It must not be possible to achieve the purpose of the research with anonymised data instead (s.251(4))

As well as these elements, the CAG's pre-application checklist adds further considerations that will be taken into account, including how patients could be informed about what is being done with their information and whether there is an 'exit strategy' and measures that may be taken to carry out the activity without using identifiable information or to seek consent from patients.³⁹

2.3.2 Regulation 3 Communicable disease and other risks to public health

Regulation 3 both *permits* the processing of confidential patient information and allows the Secretary of State to *mandate* processing, for a range of public health purposes including (under Regulation 3(1)):

- (a) diagnosing communicable diseases and other risks to public health;
- (b) recognising trends in such diseases and risks;
- (c) controlling and preventing the spread of such diseases and risks;
- (d) monitoring and managing—
 - (i) outbreaks of communicable disease;
 - (ii) incidents of exposure to communicable disease;
 - (iii) the delivery, efficacy and safety of immunisation programmes;
 - (iv) adverse reactions to vaccines and medicines;
 - (v) risks of infection acquired from food or the environment (including water supplies);
 - (vi) the giving of information to persons about the diagnosis of communicable disease and risks of acquiring such disease.

In terms of the permissive power, not everyone is empowered to process this information but a potentially wide group of professionals are, if they are: employed by the health service or engaged 'for the purposes of the health service or employed or engaged by a Government Department or other public authority in communicable disease surveillance'.⁴⁰ This allows the processing of CPI by public health organisations, such

as Public Health England, for a range of relevant activities relating to COVID-19 as well as other diseases and risks.

In terms of the power to *mandate* processing, paragraph (4) of Regulation 3 empowers the Secretary of State to require processing for these purposes by issuing a ‘notice’ which can set out which information should be processed, for which specific purpose and within what timeframe, namely the ‘COPI Notices’.⁴¹

2.3.3 The COPI notices

In March 2020, early in the COVID-19 pandemic, the Secretary of State for Health and Care issued four notices. One notice required:

- General practices
- Local authorities
- Combined authorities
- Arm’s-length bodies of the Department of Health and Social Care

to process confidential patient information to support the Secretary of State’s response to COVID-19 (a so-called ‘COVID-19 purpose’).⁴²

Another notice⁴³ required NHS England & Improvement to process confidential patient information for a COVID-19 purpose, where ‘requested to do so by an authorised officer of the Department of Health and Social Care’ acting on the Secretary of State’s behalf, or by ‘another organisation permitted to process confidential information under Regulation 3(3) of COPI (the Requestor)’. As discussed above, this is a wide group of potential requestors.

A third notice⁴⁴ was more specific requiring all ‘GP practices, whose IT systems are supplied by TPP or EMIS’ (the two major suppliers of primary care IT systems in the UK) to ‘release the relevant primary care data to UK Biobank for purposes related to the outbreak of COVID-19’. This is a slightly different description of purposes but the indicative list of such purposes in the notice is very similar to those in the first. This notice *mandates* the disclosure of confidential information that may already have been authorised by the UK Biobank participant’s explicit consent, requiring GPs to take action rather than relying on their discretion.

A fourth, and currently final notice,⁴⁵ required the Health and Social Care Information Centre, known as NHS Digital, to process confidential patient information ‘to support the Secretary of State’s response to COVID-19 (COVID-19 purpose)’. This notice is justified on the basis ‘that it can lawfully and efficiently disseminate confidential patient information’ in connection with the health and social care system’s management of the response to COVID-19. The implication of this justification is that there is either a form of dissemination of CPI that NHS Digital is legally barred from conducting, or not explicitly empowered to conduct under the normal legal framework, but it is not clear what form/purposes or recipients this barrier applies to. NHS Digital is only required to disseminate such confidential patient information where it is:

- requested to do so by an authorised officer of the Department of Health and Social Care acting on behalf of the Secretary of State or requested to do so by another organisation permitted to process confidential information under Regulation 3(3) of COPI (the Requestor) and
- reasonably satisfied that the confidential patient information to be disclosed pursuant to the request is required by the requestor for a COVID-19 purpose and will be processed by the requestor or by a processor on behalf of the requestor, solely for that COVID-19 purpose and in accordance with the restrictions set out in Regulation 7 of COPI

What happens when the COPI notices expire?

At present the COPI notices are due to expire on the 31st March 2022. However, the notices may be modified or extended by the Secretary of State prior to that deadline. If the notices expire, the addressees will no longer be required to process CPI for COVID-19 purposes. For research this means that ongoing sharing of data may cease to take place but it does not necessarily mean that the recipient and any subsequent secondary users should delete the data completely. However, the use of such data will be very limited indeed without further authorisation for disclosure.

Once the notices expire, their authorisation for disclosure of CPI will no longer apply and an alternative will need to be found for any further or subsequent disclosures. Most likely alternatives include consent or approval under Regulation 5 for research disclosures. New approval or authorisation may no longer restrict the use of CPI to COVID-19 purposes but it is feasible that the original basis for disclosure will be taken into account and researchers may be asked to limit their use of the CPI to the same COVID-19 purposes.

The principles of fairness and transparency in data protection law (see 2.4 below) may also imply that it would not be fair to change the purposes of processing this data without significant effort to communicate this to patients and the public, even if consent is not the legal basis for processing.

2.3.4 The scope and limits to processing under the COPI notices

These notices all enable and mandate processing of CPI for a 'COVID-19 purpose'. As some of the notices remind the recipient, such purposes must be within the scope of Regulation 3(1) and the broad purposes described above.

COVID-19 purposes

In two of the notices a non-exhaustive list of COVID-19 purposes is provided. These include:

- understanding COVID-19 and risks to public health, trends in COVID-19 and such risks, and controlling and preventing the spread of COVID-19 and such risks
- processing to support the NHS Test and Trace programme

- identifying and understanding information about patients or potential patients with or at risk of COVID-19, information about incidents of patient exposure to COVID-19 and the management of patients with or at risk of COVID-19 including: locating, contacting, screening, flagging and monitoring such patients and collecting information about and providing services in relation to testing, diagnosis, self-isolation, fitness to work, treatment, medical and social interventions and recovery from COVID-19
- understanding information about patient access to health services and adult social care services and the need for wider care of patients and vulnerable groups as a direct or indirect result of COVID-19 and the availability and capacity of those services or that care
- monitoring and managing the response to COVID-19 by health and social care bodies and the government including providing information to the public about COVID-19 and its effectiveness and information about capacity, medicines, equipment, supplies, services and the workforce within the health services and adult social care services
- delivering services to patients, clinicians, the health services and adult social care services workforce and the public about and in connection with COVID-19, including the provision of information, fit notes and the provision of healthcare and adult social care services
- research and planning in relation to COVID-19⁴⁶

From this list the most relevant to research is the final example of ‘research and planning in relation to COVID-19’. This has the potential to enable a very wide range of research that could be said to relate to COVID-19 and given this is simply an indicative list, the scope is potentially very broad. This raises a number of questions, including who determines whether processing is for an acceptable COVID-19 purpose and what some of the limits of such purposes may include (e.g. research in other diseases that relate to this virus or which may place a burden on health systems, thereby impacting the management of COVID-19).

Restrictions and exclusions

Regulation 7 of the COPI Regulations sets some restrictions and exclusions to the processing under the Regulations and COPI notices. These include that as minimal information as possible should be processed to achieve the permitted purposes; information should be de-identified as far as possible; access should be limited to those who are required for and aware of the purposes of processing; technical and organisational measures are taken to prevent unauthorised processing, and; the need to process CPI should be reviewed by the person in possession of that information at least every 12 months. These requirements echo some of the stipulations of data protection law.

Indeed, a key restriction on the Control of Patient Information Regulations is that they cannot modify the application of data protection law.⁴⁷ This means that the data protection legislation (UK GDPR and DPA 2018) must be applied independently and without modification to the processing of confidential patient information.

2.4 Data protection law

The current UK legal framework for data protection is derived from the EU General Data Protection Regulation which currently applies in virtually the same form as the 'UK GDPR'. This sits alongside the Data Protection Act 2018 which tailors and supplements some parts of the general regulation. The whole framework is overseen by the independent authority, the Information Commissioner's Office which provides guidance on the application of the law and is responsible for handling complaints, with the discretion to levy fines and carry out enforcement where there has been a breach of data protection. The courts will also hear claims of a breach of data protection law.

Data protection law and the common law of confidentiality will often apply in tandem to patient or medical data used in, or requested for, research. However, the way they relate to each other and interact is complex, and can give rise to significant uncertainty for researchers or data custodians. For the purposes of our research there are two key aspects for focus: the overlapping but potentially different scope of 'personal data' and CPI and the difference between consent in data protection law and the CLDC.

2.4.1 'Personal data' and Confidential Patient Information

The UK GDPR applies only to 'personal data'. This is information that relates to an identified or identifiable natural (and living) person (Art 4(1)). To determine whether data are 'personal data' a broad contextual assessment is required, considering a range of factors including the availability of other information that could be used to help identify an individual (see box below). The standard for determining whether information is identifying is whether there is a reasonable likelihood of identification taking into account all the circumstances and nature of the data at hand. In general, this has been interpreted broadly, so that de-identified data (for example, stripped of any name, date of birth and other directly identifying attributes) is likely to be 'personal data' if it is still individual-level (as opposed to aggregated) data.

Identifiability and 'personal data'

Recital 26 GDPR/UK GDPR: To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.

As outlined earlier, in the context of medical and health information, the ability to identify an individual is also an element of determining whether the information is confidential. The statutory definition of confidential patient information (NHS Act 2006, s251(11)) refers to the identity of an individual being 'ascertainable ... (i) from that information, or (ii) from that information and other information which is in the possession of, or is likely to come into the possession of, the person processing that information.'

Although there are some potential differences between this approach and the approach to identifiability under data protection law, for example the focus on the information available to the person 'processing that information' as opposed to the broad range of an interested and sufficiently determined person, they are broadly similar. However, the GDPR (UK GDPR) has potentially broadened how identifiability is interpreted in the UK in one important way - the inclusion of pseudonymised or coded data within the category of 'personal data'.

Pseudonymisation, data protection and confidentiality

Pseudonymisation is generally referred to as the process of de-identifying individual level data so that it is not possible to identify an individual without a key or code. By securely keeping the key separate from the rest of the data, the risks of unauthorised re-identification are greatly reduced but it remains possible to deliberately re-identify an individual if this is required. For example, in the genomic context this may take place when a new discovery has important implications for a previously tested individual, and where further contact is justified to provide further information or a change to management or treatment

Prior to the GDPR/UK GDPR, the interpretation of the UK Courts, tribunals and the ICO was that (in certain circumstances) coded or pseudonymised data could be sufficiently de-identified that they fall outside the scope of 'personal data', providing the key or code was not available to the party processing the data.⁴⁸

As our interviewees discussed, CPI has been largely interpreted along the same lines by data custodians in the NHS, with the ICO's previous guidance on anonymisation providing a key guide to determining identifiability.

However, the GDPR/UK GDPR has now expressly incorporated 'pseudonymisation' (Art 4(5)) and an implied category of 'data which have undergone pseudonymisation'(Recital 26) which has led to uncertainty about whether this data is always considered personal data and any consequent implications for the interpretation of CPI.⁴⁹ There is the potential that the approach to identifiability under data protection law will diverge from the approach taken in relation to CPI if the same standards continue to be applied to CPI as prior to the GDPR. From a legal perspective, there has been no change to the statutory definition of CPI, therefore the default expectation is that there will be no change to how it is interpreted.

2.4.2 Consent to research under data protection law and the CLDC

Consent, or, another form of authorization for disclosure of CPI, is the central requirement of the common law duty of confidentiality. Data protection law goes much further by setting a range of obligations and rights that apply whenever ‘personal data’ are processed. However, the UK GDPR requires all processing of personal data to have a specific legal basis from a menu of six options (Art 6(1)) and consent is one possibility. For researchers and those counselling patients/participants, it would be most straightforward if the same consent process and standards could be used to satisfy both purposes. This was more frequently the approach taken until the implementation of the GDPR but the Regulation has introduced different and higher standards for consent in data protection law that means that it has become poorly suited to some forms of genomic health and research.

There are three particular challenges. One is that the European Data Protection Board’s (EDPB) authoritative statements have suggested that there may be an imbalance of power in circumstances where the research participant is not in good health, which could mean that consent is not ‘freely given’.⁵⁰ Another, is that consent is required to be specific, and this has also been interpreted by the EDPB and its predecessor body restrictively for research.⁵¹ There is some debate in the academic literature about the potential for a valid ‘broad consent’ to research under the GDPR⁵² but the prevailing view is that research purposes need to be tightly specified at the outset for valid consent (or capable of being specified as research progresses), which can be challenging in the case of more open areas of research. Thirdly, if research is based on consent under the GDPR then a data subject is free to withdraw that consent at any point, requiring research based on that data to stop. This does not align well with common forms of consent to research where withdrawal may impact research that has already taken place and the processing of data obtained prior to withdrawal.⁵³

Taken together, these challenges for consent under the GDPR have led many data controllers and organisations to recommend against consent as a legal basis for health and social care research,⁵⁴ including the UK NHS Health Research Authority.⁵⁵ This means that researchers will often rely on an alternative legal basis (such as the performance of a task in the public interest) for processing personal data. As we noted in previous research, this may lead to difficulty communicating these differences to participants and explaining that their consent does not apply for data protection purposes.⁵⁶

2.5 Distinguishing between surveillance and research

As discussed earlier in this section, there is a considerable blurring of the boundary between public health activities, such as disease surveillance, and ‘research’. Interviewees also raised concern about the uncertainty caused by this blurring in terms of determining which information governance framework should be applied. For example (and as set out in Section 3.3), confidential patient information may be processed without consent for a wide range of public health purposes under Regulation 3 of the COPI

Regulations, including (a) diagnosing communicable diseases and other risks to public health; (b) recognising trends in such diseases and risks; (c) controlling and preventing the spread of such diseases and risks.

Distinguishing between activities such as ‘recognising trends’ in diseases and risks which fall within the public health framework and those which are part of research is difficult. They are likely to involve very similar methodology and data processing, and they are each likely to identify new knowledge which in turn will influence medical and public health practice.

The UK Policy Framework for Health and Social Care Research defines research as: ‘the attempt to derive generalisable or transferable ... new ... knowledge to answer or refine relevant questions with scientifically sound methods’⁵⁷ and it emphasises that a crucial factor is the intention of the investigators: ‘i.e. the project deliberately uses methods intended to achieve quantitative or qualitative findings that can be applied to settings or contexts other than those in which they were tested.’⁵⁸ This is also an important element in the HRA’s decision tool distinguishing research from service evaluation, clinical/non-financial audit and ‘usual practice’ - which includes public health activities.⁵⁹ This tool (and the table it draws on) demonstrates that usual public health practice may involve many similarities with research, including answering a question about health issues in a population and how they can be addressed, potential use of systematic, qualitative or quantitative methods, analysis of existing routine data and even, potentially, an element of randomisation. The crucial distinction is made on the basis of the intention, with public health practice being described as having a narrower intention such as investigating an outbreak or incident to help in disease control and prevention, as opposed to the production of generalisable or transferable new knowledge, which is the domain of research.

The challenge with this blurred boundary is that it makes a significant difference to the information governance that applies to an activity. Research requires (almost always) research ethics committee review and in most ordinary cases will be based on participant consent to the disclosure of confidential information, where the research uses identifiable patient level data. In cases where that is not feasible the research may be authorised under Regulation 5 of the COPI Regulations.

2.6 The National Data Opt-out

One further aspect of the information governance landscape that is not technically part of the regulatory framework is the National Data Opt-out. This, and the opt outs it replaces are the central means of facilitating patient choice about the use of their confidential patient information for purposes beyond direct care.

The National Data Opt-out policy was introduced in 2018 to enable patients to opt-out from the use of their data for research or planning purposes. This was recommended by the National Data Guardian who had

identified that, despite the NHS Constitution providing a right to request that personal confidential data are not used beyond an individual's care, there was no easy way for them to exercise that right.⁶⁰

The National Data Opt-out (NDO) is a policy opt-out that must be considered and applied alongside existing data protection legislation, other laws and best practice.⁶¹ Previous opt-outs were introduced in 2013 as a response to the Connecting for Health initiative. A Type 1 opt-out for personal confidential data leaving the GP practice for purposes beyond direct care and a Type 2 opt-out that prevented confidential data being disseminated by the Health and Social Care Information Centre (NHS Digital) for purposes beyond their direct care (Type 2).

The Type 1 opt-out is still active (and has been a high profile means of objecting to recent plans for the General Practice Data for Planning and Research initiative). However, the Type 2 opt-out was replaced in May 2018 when the National Data Opt-out was introduced. The key difference is that the Type 2 opt-out only applied to NHS Digital, whereas the National Data Opt-out applies to all health and care organisations in England.⁶² The pandemic has impacted the implementation of the National Data Opt-out so that the deadline for compliance has been extended in line with the COPI notices to the end of March 2022.

Scope of the National Data Opt-out (NDO)

The NDO only applies to confidential patient information (defined earlier in the section). It does not apply to uses that are considered to fall within the category of 'individual care and treatment'. This includes:

- data sharing between care settings for the care of the individual (e.g. GP to hospital)
- local clinical audit
- the summary care record and local shared records

NHS Digital also sets out further 'elements of patient care which rely on the wider processing of data, but that should also be treated as individual care' which include population screening programmes and 'risk stratification used for case-finding' when carried out by the individual's care provider.⁶³

By contrast, research and planning purposes include understanding outcomes of patient care and using data to make resource and funding decisions, as well as the broad range of potential research questions that could apply to this data.

Restrictions and exemptions

The NDO does not apply to all disclosures of CPI beyond individual care. There are a range of exemptions. First, it will not apply if the patient has given explicit consent for the specific disclosure at hand (i.e. for a specific research project). Second, it will not apply where disclosure is required for the purposes within Regulation 3 of the Health Service (Control of Patient Information) Regulations 2002. As discussed above,

this includes public health activities and research and other forms of processing in line with Regulation 3 and the subsequent COPI notices.

Third, the Opt-out will not apply where information is disclosed in compliance with a legal obligation. Perhaps the most relevant of these is the power that NHS Digital has to collect information when directed to do so by the Secretary of State or NHS England under s259 Health and Social Care Act 2012.⁶⁴ The policy is also not applied to data flows to Public Health England National Diseases Registers.⁶⁵ Finally, the Opt-out does not apply where there is an overriding public interest (e.g. reporting of gun or knife wounds).

In some ways it is simpler to identify when the NDO does apply in the research context. It will apply to disclosure of CPI for research purposes under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002. As outlined above, this involves advice from the Confidentiality Advisory Group to the Health Research Authority in order to decide whether disclosure of CPI may go ahead for research purposes without consent. However, according to the National Data Opt-out Operational Policy Guidance Document, '[i]n exceptional circumstances, and on a case-by-case basis only, CAG may advise the decision-maker that the national data opt-out should not apply to a specific data flow supported under S.251'.⁶⁶

Role and limitations of the National Data Opt-out

As at 1st September 2021 there were 3,264,327 NDOs registered (5.35% of the population registered with a GP Practice), an increase of 58,304 compared to 1 August 2021.⁶⁷ This figure is likely to have been influenced by coverage of the GP data for planning and research initiatives. Although the Opt-out is a policy measure not a legal provision, it has a significant impact on disclosure of CPI for research purposes. However, it does not impact research enabled by the COPI notices or public health activities carried out under Regulation 3 (as outlined above).

As a means of giving individuals a choice over the processing of their CPI, the NDO is a relatively blunt instrument. For example, it does not allow individuals to opt out of processing by certain types of actors, such as private sector organisations, or to opt out for more granular forms of CPI as opposed to all their health information. It also does not apply to disclosure of anonymised information.

2.7 Section summary

There is an extensive and complex legal framework governing the use of confidential patient information (CPI) for genomic and medical research in England. The backbone is the common law duty of confidentiality which requires either consent (express or implied) to authorise disclosure, a statutory mechanism that sets aside confidentiality for specific purposes or a determination that disclosure is in the public interest. The Health Service (Control of Patient Information or 'COPI') Regulations provide several key routes for lawful

processing of CPI without consent, both as a statutory gateway and an indication that such processing is in the public interest. Prior to the pandemic, the route for research was under Regulation 5, which delegates decision making to the Health Research Authority, taking advice from the Confidentiality Advisory Group. This requires researchers to demonstrate that it is not possible to carry out the research in another way, including with anonymised data, and that the proposal is in the public interest.

During the pandemic, the Secretary of State for Health and Care has issued four COPI notices under Regulation 3 of the COPI Regulations on the basis that such data processing is necessary. These COPI notices enabled and mandated the disclosure of CPI by a range of key actors, without consent for COVID-19 purposes. These purposes are open ended but include research purposes providing they relate to COVID-19. Since March 2020, the COPI notices have been extended three times, most recently to until 31st March 2022.

There are some limits to processing enabled under the COPI notices, including that data protection law will continue to apply in unmodified form. This is important because the interaction between data protection law and the COPI regulations/notices is one of the issues highlighted in this section and by our interviewees as requiring further scrutiny. This is for two main reasons. First, the scope of CPI and 'personal data' is similar in terms of identifying an individual but there is uncertainty about how they may now diverge in light of changes to data protection law. This may mean that some data are considered to be 'personal data' but not confidential patient information. Second, updates to data protection law have also meant that relying on consent as a legal basis for data processing is more challenging for research. A consent which satisfies ethical requirements or to authorise disclosure of CPI will not necessarily be sufficient for data protection purposes. These are challenges for researchers, data custodians and patients/participants in understanding and explaining when or how the law applies.

Another key challenge is the blurred boundary between research and other activities, such as surveillance. This is important because different legal requirements apply depending which side of the boundary an activity falls. This could become increasingly important if the framework for research of public health activities is amended in the future.

Finally, an extensive range of guidance, policy and governance applies to the processing of CPI for research purposes in England. This has not been the focus of this evaluation but some key parts of that framework have a bearing on the legal framework and how it is applied. Most important of these is the National Data Opt-out policy, which provides an opt out to processing for purposes beyond direct care in certain circumstances. We return to consider the strengths and weaknesses of this policy and the other parts of the framework highlighted in this section, when we consider if changes to the regulatory framework made during COVID-19 should be integrated on a permanent basis. In the next section we consider how confidential patient information and other relevant genomic and health data were used for genomic and medical research during the pandemic.

3. The landscape of data use for genomic and medical research during COVID-19

3.1 Introduction

Over the last decade, genomic technologies have moved from research, through population based programmes such as the 100,000 Genomes Project, to become an integral and routine part of healthcare. The National Test Directory prescribes the genetic and genomic tests which can be utilised to support disease diagnosis, treatment and management. Therefore it is no surprise that genomic technologies are being employed during the pandemic in a research setting for a multitude of applications, some of which spill over into clinical care. The sequencing of SARS-CoV-2, the virus that causes COVID-19, from patient samples provides important information about how the virus is changing (i.e. accumulating mutations) as it spreads through the population. Genomic information is also being sequenced from humans who have and have not been infected with SARS-CoV-2 to reveal genetic influences of susceptibility and disease severity.

These data from the SARS-CoV-2 virus, from patients and from population sampling are using genomic information to understand the virus, how it infects and causes disease, how it is transmitted, why there are variable responses between individuals and if mutations in the virus genome are impacting its ability to cause disease. There is a particular focus on whether variants of SARS-CoV-2 have evolved that are a cause for concern to public health. This could be because they possess mutations that enhance the ability of the virus to transmit between individuals, alter the disease mechanisms so that more severe disease results or enable it to evade host immunity.

Genomic data, either viral or human, has limited value on its own and requires linkage to other datasets to derive scientific and clinical insights. Linkage of clinical and epidemiological datasets has been essential to answer important questions pertinent to the pandemic. Some of this data will include confidential patient information (CPI). In this section we outline the health data landscape providing access to and processing confidential patient information for COVID-19 purposes in England.

3.2 The health data landscape

There are many organisations in England that retain and process CPI including primary, secondary and tertiary care organisations. The primary purpose for health providers to hold CPI is for the direct care of patients. Data are also collected and processed for public health purposes, such as for disease monitoring and surveillance. As described in section 2, the distinction between research and surveillance is blurred with the emphasis being on surveillance at the outset of the pandemic, moving to research as the immediate public health emergency has abated. This section outlines some of the key organisations involved in collecting, processing and facilitating access to that data by researchers. These initiatives

straddle efforts to mandate the systematic sharing and collection of health data, as well as the provision of facilities to provide improved research access to data in a secure environment. This section is based on publicly available sources, supplemented by insights from our stakeholder interviews, during Spring/Summer 2021. We have updated elements to reflect key recent changes to the landscape, including the re-distribution of Public Health England's responsibilities to other agencies. Although this is an extensive picture of the health data landscape during COVID-19 in England, it is non-exhaustive and limited to information made available during our research.

Organisations/initiatives providing access to data in England

NHS Digital

Primary, secondary and tertiary care organisations each hold confidential patient data including patient's name, date of birth, address, contact details, details of health conditions and illnesses, medications and treatments being received and the contact patients have had with doctors and other health care workers. NHS Digital receives some of these data from health and care organisations in England.

NHS Digital has two main responsibilities:

- To run and manage computer systems that link different parts of health and care together
- To collect some specific health and care data to monitor the performance of the health and care service and to improve care

To carry out these duties, NHS Digital need to collect, store, process and sometimes share information about patients. NHS Digital does not hold full patient records but they collect some data on everyone. These data are linked to an individual's NHS number, a unique identifier ensuring that data is linked to the correct patient.

NHS Digital uses confidential patient information to improve individual care as well as improving the running of health services and allocation of resources. They also share confidential patient data with other organisations, usually in pseudonymised form.⁶⁸ Sometimes they allow the use of identifiable data for the purposes of improving health and care under strict control.⁶⁹ Clinicians, researchers, commissioners and other organisations can access this data via the Data Access Request Service.⁷⁰

Multiple datasets⁷¹ are available via the NHS digital Data Access Request Service including datasets specifically relating to the COVID-19 pandemic including the General Practice Extraction Service (GPES) data for pandemic planning and research (formerly GPDPR) see below. During the pandemic, NHS Digital has also received additional datasets from Public Health England (PHE) (now the UK Health Security Agency (UKHSA) and Office for Health Improvement and Disparities) under a data provision notice.⁷²

These include:

- PHE Second Generation Surveillance System (SGSS) data. This includes demographic and diagnostic information from laboratory test reports for patients tested for the suspected and confirmed causative agent for COVID-19.⁷³
- PHE COVID-19 Hospitalisations in England Surveillance System (CHESS) data. This includes demographic, risk factor, treatment, and outcome information for patients admitted to hospital with a confirmed COVID-19 diagnosis.⁷⁴

The Data Provision notice clarifies that the National Data Opt-out does not apply since these data are for direct care purposes. By implication these are not for research, although their findings might inform research.

General Practice Extraction Service

One of the most significant changes during the pandemic has been the collection and onward sharing of GP data. NHS Digital has been directed to collect and analyse health care data for the duration of the pandemic. GPs and other health organisations are now legally obliged to share patient data, including identifiable patient data, with NHS Digital for COVID-19 purposes. Before the pandemic this data sharing was voluntary and contained primarily anonymous data. Since the COPI notices, requests for access to GP healthcare data are directed to NHS Digital, as the national safe haven for health and social care data in England, rather than individual GP practices. As we noted in section 2, NHS Digital will be obliged to share data for COVID-19 purposes where requested to do so in accordance with the relevant COPI notice.

NHS Digital collected GP data fortnightly during the pandemic to support planning and research into COVID-19. All requests for access to the data are currently made through the NHSX Single Point of Contact for COVID-19.

NHS Digital Trusted Research Environment

Optimising researcher access to standardised datasets within a privacy protecting and technically secure environment has been facilitated by the NHS Digital's Trusted Research Environment (TRE) service for England. This provides approved researchers with access to linked, de-identified health data to address COVID-19 related research questions. The aim of this partnership is to provide the tools and data needed to support researchers from UK universities and other organisations in analysing a variety of linked data sources. Working in partnership with Health Data Research UK (HDR UK), NHS Digital has collaborated with partners such as the British Heart Foundation Data Science Centre (BHF DSC) to provide a secure environment for accessing data that can answer complex research questions. For example, this BHF team is currently researching the impact and effects of the COVID-19 pandemic on cardiovascular diseases in terms of diagnosis, management and patient outcomes.

The use of a TRE service also provides harmonised approaches to analysis and interrogation. In this case, TRE service users are given access to NHS Digital's Data Platform which currently hosts analysis and interrogation tools such as Databricks and RStudio. (Databricks is a collaborative analytics platform that supports SQL and Python languages for the analysis of big data in the cloud; RStudio is a data analysis environment for R, a programming language for statistical computing and graphics).

Researchers with the same data sharing agreement can work collaboratively with their colleagues in shared project folders, using their preferred tool. The final intended output is checked for compliance with the five safes⁷⁵ (Safe People; Safe projects; Safe settings; Safe outputs and Safe data) before exports are approved. By facilitating a privacy enhancing and technically secure environment for data processing, the TRE service can be used to help guide national decision making and recommend potential interventions to reduce the severity of COVID-19 outcomes, while safeguarding patient confidentiality and data protection.

NHS England COVID-19 Data Store

Another significant initiative is the secure NHS England COVID-19 Data Store which has been established by NHS England and NHS Improvement working with NHSX.⁷⁶ This Store brings together the accurate, real-time information necessary to inform decisions in response to the pandemic in England. This includes data already collected by NHS England, NHS Improvement, Public Health England and NHS Digital.

This enables the NHS and the government to monitor the spread of the virus, identify trends and implement appropriate measures to ensure services and support are available to patients, e.g. to analyse bed capacity in hospitals or the provision of ventilators in a particular area.

Datasets provided by NHS Digital are pseudonymised prior to going into the NHS Data Store to ensure that individual patients are not identifiable. However datasets from PHE and the Intensive Care National Audit and Research Centre are received in identifiable form and are pseudonymised by NHS England. These include the following datasets:

- Identifiable data (including laboratory test data) from Public Health England via NHS England
- Data from the COVID-19 Hospitalisation in England Surveillance System (CHESS) database
- Data concerning the care and discharge data of COVID-19 patients from the Intensive Care National Audit and Research Centre (ICNARC)

A single point of contact (also referred to as a single front door) has been established to manage requests for access health and care data held by NHS England and NHS Improvement, NHS Digital or Public Health England in order to support the COVID-19 response.⁷⁷ A data dissemination register demonstrates how some data are being shared for COVID-19 purposes through this process.

The National Immunisation Management Service

NHS England has also established a centralised service for the management of both the COVID-19 and seasonal flu vaccination programmes. This service is supported by a central system, the Immunisation Management System.⁷⁸ The purpose of this system is to enable identification of priority groups, to send invitations to book appointments for vaccination, to manage and monitor the progress of the programme. Data from the National Immunisation Management Service is available on the NHS England COVID-19 Data Store. In addition, personal data from the Immunisation Management System is shared with the following external agencies:

- Public Health England (PHE) – an executive agency sponsored by the Department of Health and Social Care
- Joint Biosecurity Centre (JBC) – a directorate of the Department of Health and Social Care
- Trusted Research Environments – operated by a number of organisations including the Office for National Statistics (ONS)
- SPI-M – an independent group set up by the Government to support the Scientific Advisory Group for Emergencies (SAGE)
- NHS Digital – joint controller with NHS England for processing to facilitate the analysis, linkage and dissemination of data about COVID-19 vaccination (under the COVID-19 Public Health (NHS England) Directions 2020) to requestors who have an appropriate legal basis to process it.

Coronavirus (COVID-19) Research Platform

NHS England processes CPI to identify medical conditions and medications that affect the risk or impact of COVID-19 infection on individuals; this will assist with identifying risk factors associated with poor patient outcomes as well as generating information to monitor and predict demand on health services. They receive data from the following sources:

- COVID-19 Hospitalisation in England Surveillance System (CHESS) (Public Health England), Intensive Care National Audit and Research Centre (ICNARC) and other NHS intensive care or relevant datasets containing information about the healthcare of patients with COVID-19;
- Primary care (GP) records processed by TPP (i.e. GP surgeries using SystmOne software), one of the GP electronic health record providers.

COVID-19 National Core Studies

The UK Government has established six complementary National Core Studies⁷⁹ which have the objective of using health data and research to inform both the near and long-term responses to COVID-19, as well as accelerating progress to establish a world-leading health data and research infrastructure for the future. Notably, the Data and Connectivity National Core Study⁸⁰ works to make vital data available to accelerate

research on COVID-19, and is led by HDR UK in partnership with the Office for National Statistics. The other five core studies have focused on different aspects of COVID-19 and are led by other stakeholders, with data access facilitated via TREs (including NHS Digital).

Health Data Research UK

Health Data Research UK (HDR UK) is an independent non-profit organisation supported by thirteen charities and public bodies with a mission to unite health and care data across the UK to enable discoveries that improve people's lives.

HDR UK is contributing to the pandemic response in several ways through strategic alliances with key stakeholders. Firstly, it is providing the Strategic Advisory Group for Emergencies (SAGE) fortnightly updates⁸¹ regarding prioritised health data research related to COVID-19. In addition, through provision of a gateway, HDR UK facilitates linkage of COG-UK COVID-19 genome databases with epidemiological and clinical databases nationwide.

HDR UK's existing national network of patients and members of the public, including members of the HDR UK Public Advisory Board and HDR Hub Public & Patient Advisory Groups participate in deciding which research questions are most important and urgent to warrant access to datasets including those held by NHS Digital.⁸²

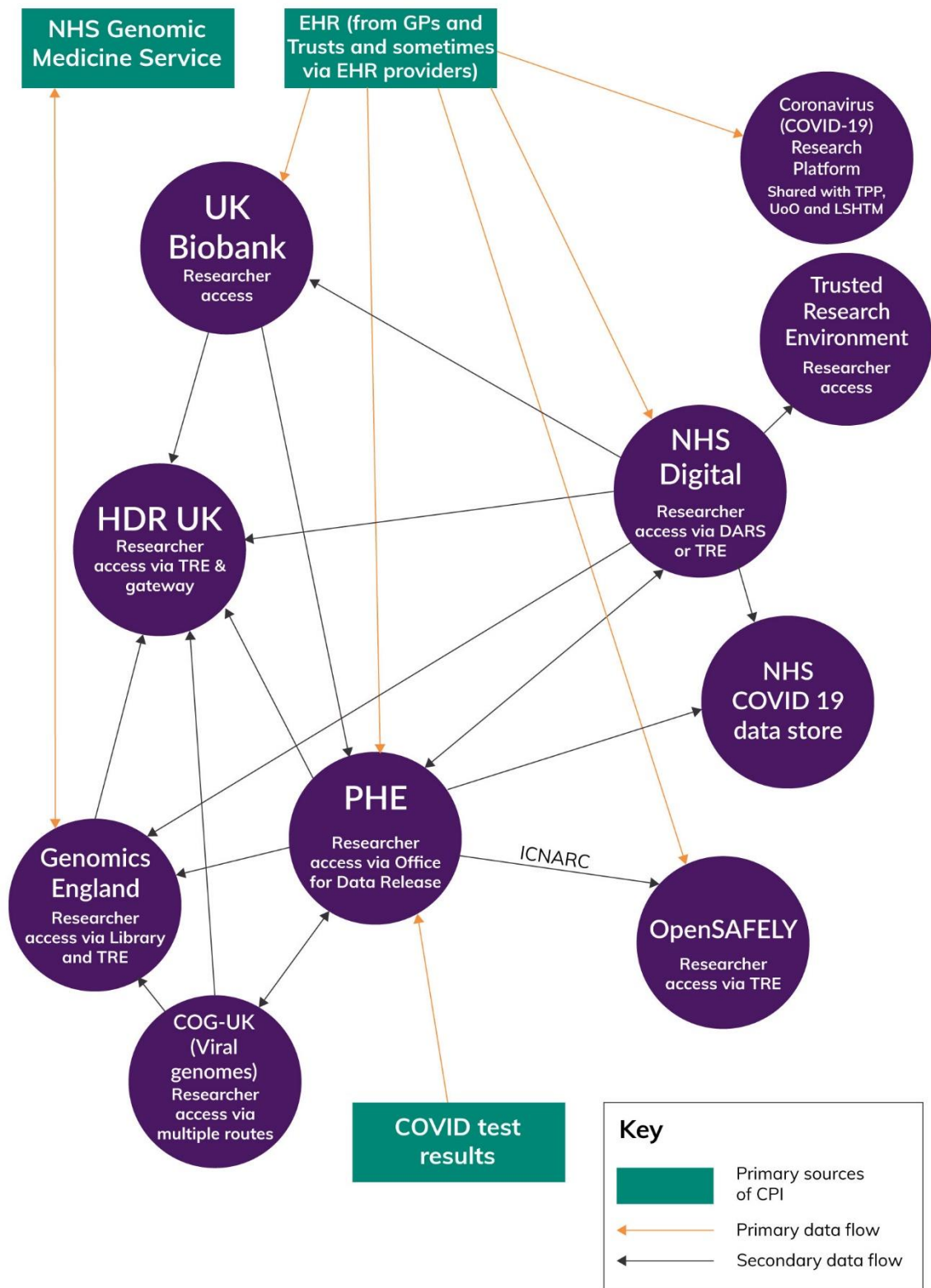
Health Data Research UK Innovation Gateway

The HDR UK Innovation Gateway has had an important leadership role in the Data and Connectivity National Core Study⁸³ in collaboration with Trusted Research Environments. The Innovation Gateway provides a common entry point for researchers to access the data sets held in the NHS Digital TRE⁸⁴ service as well as many additional health research data sets held in thousands of disparate organisations around the UK. It provides detailed descriptions of the datasets available, facilitates access and provides advanced research tools.

One notable dataset in the context of genomic research contains over 200,000 SARS-CoV-2 viral genome sequences published by COG-UK. Linking existing data collections has produced a national registry of COVID-19 patients (infected and recovered) for research. All these data are combined with the viral sequences in COG-UK's central database, called CLIMB-COVID-19. The data is de-personalised and can be made available as open access to any COVID-19 researchers.⁸⁵

Figure 1 summarises the key data sources and data flows described in this section. This diagram is non-exhaustive and is intended to demonstrate the complexity of the data flow landscape through mapping out some of the key data flows. This describes data flows which were active in June 2021.

Figure 1: Diagram of key data flows for research using confidential patient information (CPI) during COVID-19



Summary

There have been multiple efforts to collect and aggregate data sources which could inform government and health system responses to the COVID-19 pandemic. Many of these efforts have been by government agencies, to coordinate, streamline and mandate data collection. HDR UK has also played a key role in facilitating researcher access to these diverse datasets in collaboration with TRE providers. The following section explores in more detail how genetic and genomic data have been generated and utilised during the COVID-19 pandemic.

3.3 Genomic surveillance

Confidential patient data, de-identified or anonymised data have been analysed and evaluated for two potential purposes:

- the surveillance of individuals and populations over time to understand trends in disease, to make predictions about the future and to inform management and treatment of individuals and populations
- research to provide deeper understanding through generating novel, and potentially generalisable findings

As the pandemic has progressed, both surveillance and research activities have increased considerably compared to pre-pandemic levels. These activities may utilise the same data but interrogate these differently through different research questions and methodologies. Some important initiatives such as the COVID-19 Genomics UK Consortium sit on the cusp of research and public health activities. In view of the difficulties in distinguishing between these two, our analysis includes examples of genomic surveillance.

Viral genome data provides important information about how the virus is mutating over time and space. Linking viral genomic data with clinical and epidemiological data from patients allows researchers and public health agencies to understand the impact of viral mutation on transmission, disease severity and immune escape to inform national public health planning.

COVID-19 Genomics UK Consortium

The COVID-19 Genomics UK (COG-UK) Consortium⁸⁶ was the first initiative that has enabled large-scale genomic epidemiology to guide and inform the public health response to a pandemic in the UK. The initiative was established in April 2020 supported by £20 million funding from the COVID-19 rapid-research-response “fighting fund” from the Government administered by the National Institute for Health Research (NIHR), UK Research and Innovation (UKRI), and the Wellcome Sanger Institute.

It includes multiple academic universities, institutions, Public Health groups and NHS organisations, the Wellcome Sanger Genome Institute (the central genome sequencing hub), the four public health agencies of the UK, and the Lighthouse Labs. COG-UK’s original goal was to sequence positive SARS-CoV-2 samples

from up to 230 000 patients, health-care workers, and other essential workers in the UK with COVID-19. This is to enable the tracking of SARS-CoV-2 transmission, identify viral mutations, and integrate with health data to assess how the viral genome interacts with cofactors and to understand the consequences of COVID-19.⁸⁷

COG-UK sits at the interface between public health action and academic research. It shared data with both public health agencies directly and to the academic community via multiple research databases and web applications.

Viral genome sequencing data has been integrated within UKHSA and NHS Test and Trace broader surveillance, to help understand outbreaks and strengthen infection control measures across the country. Public health agencies merge genome data with detailed epidemiological and clinical data. It is this combination of data that allows them to interpret the significance of mutations for human health. COG-UK does not have access to this detailed patient level data. For the most part of the pandemic it has not been possible to sequence all positive samples therefore COG-UK developed a sampling strategy to concurrently enable broad population-level analyses, targeted analyses of specific populations, and freedom to tackle local priorities. COG-UK publish a sequencing Coverage Report each week, which is cascaded to the public health agencies, and provides details of the random sequencing of positive samples across the UK. A short version is open access and widely available.⁸⁸ As a companion to the coverage report, COG-UK also releases a Summary Mutation Report.⁸⁹

In addition to these population or targeted surveillance efforts, COG-UK collaborates on a number of research projects (e.g. the HOCl study). Outputs generated for public health purposes such as the Coverage and Summary Mutation Reports are used by other organisations and groups to assess the possible biological significance of the mutations, and decide which to prioritise for rapid investigation in laboratory studies of virus behaviour and immunology. COG-UK data has enabled a burgeoning number of publications.⁹⁰

COG-UK has recently divested routine genomic sequencing to public health authorities,⁹¹ and aims to enhance its sequence data by additional data integration with other databases, and to strengthen international collaborations, thus enhancing its research capacity.⁹²

Use of personal data

Each sample submitted to COG-UK for sequencing is given a unique identifier, which is used to link sequences and associated data together across the consortium. Diagnostic testing laboratories give COG-UK information about when and where a sample was collected as well as non-identifying information about the person who took the test including: the first part of their postcode, their age and sex. Hospital labs may also be able to give information about whether the person is a patient, member of staff or a care home resident. Public health authorities can also provide COG-UK with further data about individuals whose viral

samples have been tested without giving COG-UK identifiable data. Although these data are coded and pseudonymised, they may be considered 'personal data' within the legal framework.

Linking viral genomic data with other routine data can help researchers answer the following critical questions:

- How do SARS-CoV-2 mutations impact on the severity of disease, the transmission of the disease, and the outcomes including risk of "Long COVID"?
- Is there any interaction between viral mutations and human genomics that influence severity and outcomes?
- How do different treatments impact on the clinical disease associated with different SARS-CoV-2 variants?
- How do different variants spread in different groups of people?⁹³

Public Health England

During the pandemic, Public Health England (PHE) (now UKHSA)⁹⁴ linked the viral sequences produced by COG-UK to the other records they hold about people diagnosed with COVID-19. This allowed them to use genomics in the public health response. Accessing information about the individuals the sample came from is essential for understanding the impact of viral mutation. PHE are using a variety of data sources⁹⁵ for COVID-19 surveillance relating to: confirmed cases, community surveillance, primary and secondary care surveillance data, mortality surveillance, seroprevalence surveillance and international situation monitoring via online sources and WHO reports.

The Second Generation Surveillance System

The Second Generation Surveillance System (SGSS) is the national laboratory reporting system used in England to capture routine laboratory data on infectious diseases and antimicrobial resistance. PHE required all diagnostic testing laboratories to report all positive cases of notifiable diseases including identifiable patient data.⁹⁶ This includes mandatory reporting of tests for certain infectious diseases pursuant to the Health Protection (Notification) Regulations (2010). PHE linked data from the SGSS to UK Biobank participant data, which can be accessed by researchers through the usual UK Biobank application process.⁹⁷

COVID-19 Hospitalisation in England Surveillance System

As mentioned above, the COVID-19 Hospitalisation in England Surveillance System (CHESS)⁹⁸ collects epidemiological data (demographics, risk factors, clinical information on severity, and outcome) on COVID-19 infection in individuals that have been hospitalised and those who were admitted to ICU/HDU. This surveillance system, which was adapted from the UK Severe Influenza Surveillance System (USISS) is helping

to monitor the impact of severe COVID-19 infection on the population and health services, and provide real-time data to forecast and estimate disease burden and health services utilisation.

Linking COG-UK data

Data about individuals' from SGSS and CHESS have been linked to the viral sequence data from their sample.⁹⁹ This is essential to determine if SARS-CoV-2 variants have concerning epidemiological, immunological or pathogenic properties. Linking variant data to individual patients helps to show association between specific pathogen variants, A&E visits, hospital admissions and deaths. If suspected of being a concern, variants are designated Variant Under Investigation (VUI) status with a year, month, and number. Following a risk assessment by experts, they may be designated Variant of Concern (VOC). For example, PHE, with Imperial College, The University of Edinburgh, The University of Birmingham and the Wellcome Sanger Institute have a Novel Variant Incident Management team that carry out risk assessments for VUIs and VOCs.¹⁰⁰

Public health agencies are central to evaluating whether the roll-out of vaccination will lead to selection for mutations that allow the virus to escape from the effect of the vaccine. Part of their role is to rapidly detect individuals who have had infection more than once, or have had the vaccine but have become infected, as mutations in the virus could be driving these infections. These cases need to be prioritised to have their virus sequenced by COG-UK. This depends on effective and rapid triage by the diagnostic testing system before their test is processed to allow for efficient and comprehensive capture of cases that require further testing by sequencing. PHE and others investigated the effectiveness of vaccines against VOCs. To do this they linked COVID-19 test results, viral genomic data and vaccination status from the national vaccination register (the National Immunisation Management System, NIMS). This study noted its reliance on the COPI notices.

New and Emerging Respiratory Virus Threats Advisory Group

The New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) is an expert committee of the Department of Health and Social Care (DHSC) that advises the Chief Medical Officer (CMO) and, through the CMO, ministers, DHSC and other government departments. It provides scientific risk assessment and mitigation advice on the threat posed by new and emerging respiratory viruses and on options for their management.

PHE uses integrated datasets to determine the impact of variants on transmission and clinical outcomes. For example, initial assessment by PHE's New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) of disease severity through a matched case-control study reported no significant difference in the risk of hospitalisation or death in people infected with confirmed B.1.1.7 infection versus infection with other variants.¹⁰¹

Summary

There have been concerted efforts to establish surveillance of SARS-CoV-2 transmission at local, national and international levels. Characterising the impact of novel strains relies upon extensive data integration. Translating this knowledge into a public health tool which can be used for real-time tracking typically requires more extensive research and genomic sequencing capability.

3.4 Genomic research reliant on COPI notices

Genomic data from viral samples and patients are revealing important insights into how SARS-CoV-2 causes disease, how it spreads between individuals, who is at higher risk of severe disease, if some people are more susceptible than others and what the impact of viral mutation is on any of these factors. Examples of research taking place that is using confidential patient information linked to genomic information about the virus, patients or that is investigating interactions between virus, patient genomes and clinical outcomes are described below. Some or all of the data linkages in these studies rely on COPI notices.

Table 2: Research studies relying on COPI notices (see Appendix 2)

Name of Study	Details	Data Linkage	Potential Insights	Published Findings
GenOMICC ¹⁰² Appendix 2(1)	The role of genetics in risk of developing severe COVID-19 [20,000 severe cases 15,000 controls]	Patient genome data, clinical data (from multiple datasets including NHS Digital, IGNARC, viral genomes)	The influence of genetics independent of other risk factors; virus-host interactions	Interim findings showed a number of regions associated with severe disease. Several drug targets and treatments identified
Hospital-Onset COVID-19 Infections Study (HOIC) ¹⁰³ Appendix 2(2)	Phase III prospective interventional cohort study to evaluate the benefit of rapid COVID-19 sequencing on infection control [2000 cases with hospital onset infections]	Viral genomic data, patient data and ward level data	Impact of rapid sequencing and report production (48 hours after sampling vs 5-10 days) on occurrence and transmission. Inform future decisions to utilise pathogen genomic sequencing	Early proof of concept study published. Ongoing work to provide more immediate feedback to support public health activities
Sequencing and Tracking of Phylogeny in COVID-19 ¹⁰⁴ Appendix 2(3)	To identify distinct clusters of viral genomes to understand trends in virus adaptation	Viral genomes with 'anonymous' patient information	By analysing the nature and speed of changes in viral genomic data, can identify potential transmission and compare global	Online dashboard displaying SARS-CoV-2 lineages

			spread, identify viral spread, and assess impact of viral strains on patient outcomes	
SIREN study ¹⁰⁵ Appendix 2(4)	Determine incidence, characteristics and potential of new infections in NHS workers [44,546 workers across 135 sites]	Questionnaire responses, COVID-19 PCR tests, antibody tests, clinical and vaccination records. Positive samples are sent for viral genomic sequencing	Whether prior infection with SARS-CoV2 protects against future infection; risk assessment; surveillance of potential threats from new variants	Preliminary findings showed that a previous history of SARS-CoV-2 was associated with 83% lower risk of infection ¹⁰⁶
UK Biobank ¹⁰⁷ Appendix 2(5)	Lifetime provision of biological samples and health data [500,000 people aged 40-69 years at recruitment]	Linkage of COVID-19 test results with health records including primary care data and lifestyle data	Allows identification of individuals with COVID-19 across full spectrum of severity facilitating greater understanding of the impact of genetics, lifestyle and health on outcomes, severity and recovery from COVID-19	Many active studies and publications related to COVID-19. These include assessment of risks of severe COVID-19 and association with other diseases/genotypes ¹⁰⁸

Taken together, these studies illustrate a range of study methodologies, but they all share reliance on COPI notices to facilitate researchers' access to data and/or to linkage of different datasets. Typically, this involves the linkage of genome data, either viral or patient genome data or both, with clinical data of various types. Sometimes this clinical data is limited in time (e.g. exposure to an infection risk); in other cases these health and clinical data are much more wide-ranging and cover primary and secondary care data from multiple sources together with lifestyle data (as from the UK Biobank participants). Given that all the studies listed above explicitly state that they rely on COPI notices to facilitate their access to data and or linkage of different datasets, they illustrate the significant potential breadth and impact that the COPI notices have had on research on SARS-CoV-2 and COVID-19.

These findings have been reinforced by our interviews, many of which have highlighted that the COPI notices have facilitated research which would have been more difficult to establish prior to the pandemic. Reliance on the COPI notices have potentially served a dual function: they have mandated and streamlined

the flow of confidential patient data for which there was an existing legal basis, as well as providing access to confidential patient information for which no clear previous legal basis existed.

3.5 The future of COVID-19 research and surveillance

Scientists consider that COVID-19 is likely to become endemic.¹⁰⁹ As long as the virus continues to be transmitted between individuals, new variants of SARS-CoV-2 will continually arise. Some of these variants may continue to pose a threat to UK citizens. Interpreting viral genomic data relies on access to epidemiological datasets including patient data. This data linkage provides important contextual information as well as indicating whether specific variants are causing increased transmission between individuals. Pathogen genome data alongside patient records can also elucidate cases of repeated infection/ infection after vaccination, both are important indicators of the virus developing immune escape mutations, and if variants are causing more severe disease.

As this is a novel virus there is not a clear boundary between surveillance and research with the two heavily influencing and informing each other. This has implications for how existing research enabled or facilitated through the COPI notices may continue, and the requirements that are put in place for future research.

Viral mutation and the impact on public health

SARS-CoV-2, like all viruses, accumulates mutations – changes in its genetic code – over time as it replicates. SARS-CoV-2 evolves at a rate of $\sim 1.1 \times 10^{-3}$ substitutions per site per year, corresponding to one substitution every ~ 11 days.¹¹⁰ This is slower than some other viruses such as HIV that display $\sim 4 \times 10^{-3}$ substitutions per site per year.¹¹¹ Over the course of the pandemic the virus has diverged into thousands of different genetic variants. The mutations detected in each SARS-CoV-2 sample serve as a useful ‘barcode’ for tracking viral spread and evolution.

Mutations accumulate sporadically, however, what determines if they stay in the gene pool is if they have any biological impact on the virus and/or how it interacts with its host. The vast majority of mutations that accumulate will be neutral (i.e. they have no impact on viral or disease biology). However, mutations that confer a positive impact on the virus in terms of its ability to be transmitted to more individuals, infect its host and/ or evade host immunity will be selected for (i.e. the mutation will continue to be present in future generations). On the other hand, if a mutation is detrimental to the virus survival then it will reach an evolutionary dead end.

Determining impact on viral biology

Transmissibility – To determine statistically if there is a meaningful difference in transmission between variants, scientists ideally need to observe repeated independent introductions of each variant into the same population and follow the trajectories of the outbreaks they cause.¹¹² Frequency of VUI/VOC may also be monitored by gene target failures during PCR testing (e.g. S-gene target failure for the B.1.1.7 variant).¹¹³

Genomic surveillance can indicate if infection rates in specific geographical areas where a particular variant is circulating have increased faster than expected despite control measures being in place. Computational modelling of cases over time can therefore demonstrate whether a variant has a higher transmission rate than other variants in circulation.

Disease severity – Linking pathogen genome sequence data with clinical data on patient outcomes (e.g. patient outcome after 28 days post-diagnosis (death or recovery)) can determine whether specific variants are associated with an increased risk of severe disease.

Immune evasion – There are several ways to determine if viral mutation has an impact on host immune response. Antibody neutralisation studies test the ability of antibodies from individuals who have previously been infected or have received immunisation for SARS-CoV-2 to neutralise (bind to) the virus.¹¹⁴ Sequencing the genome of virus samples from patients with repeated infections or infection after vaccination can elucidate if specific variants are associated with host immune evasion.

All of these impacts of viral mutation require follow up with laboratory experiments to characterise the biological mechanisms leading to the observed changes. Confidential patient information allows impacts on disease severity and the ability to evade the immune system to be assessed in more detail.

The end of the pandemic?

The need for ongoing surveillance of new mutations provides a continuing need for oversight of the SARS-CoV-2 virus. One of the questions raised as a result of this project is to question at what point the need for surveillance of new mutations in the SARS-CoV-2 virus and monitoring associated COVID-19 outbreaks might cease to be a justification for the regulatory measures that have been put in place. The World Health Organisation's guidance on pandemic preparedness and response describes a 'post-pandemic period' where levels of disease return to normal 'seasonal' levels in those countries with adequate surveillance.¹¹⁵ Determining when the pandemic might be at an 'end' is not straightforward, since the pandemic is a worldwide phenomenon and successive peaks and troughs in levels of infection have been observed since the virus was first reported. However, there are precedents for moving from an emergency response to a novel pandemic, to 'surveillance' in the case of some influenza subtypes (HINI) and moving towards this type of model seems likely in this case.

In other pandemics, the nature and the scale of the surveillance that has been required depends in part on the mutability of the virus in question. For example, human immunodeficiency virus (HIV) typically mutates around four times faster than the SARS-CoV-2 virus and needs closer surveillance in order to track emerging novel variants. Other relevant factors which might influence the nature and type of surveillance required are the prevailing selection pressures for the virus e.g. rising rates of natural immunity within the population and vaccination rates. Since novel viral mutations might be more likely to arise in vulnerable populations (such as people who are immunocompromised), maintaining the capability to integrate

multiple forms of data such as clinical data sets, geographical data and viral genome data for these clinically vulnerable sub-populations might also be an important tool in early detection of novel outbreaks. Some diseases, such as seasonal influenza, can cross species barriers, enabling animals to be an additional source of novel mutations or a vector for disease transmission. To date this is not common with SARS-CoV-2, but this needs to be kept under review.

In the light of these scientific considerations, a key question is whether the COPI notices would, in future, continue to be an effective and proportionate mechanism to promote and optimise the generation, collection and sharing of confidential patient data, or whether other mechanisms might allow sufficient data utilisation whilst safeguarding public trust and confidence.

3.6 Section summary

The pandemic has exemplified how important data is for understanding and responding to a global crisis. Medical research using patient data has proven to be essential for understanding and characterising SARS-CoV-2 and COVID-19, the disease it causes. Information about patients who have been infected with SARS-CoV-2 is essential for understanding how host genomic factors influence disease susceptibility and outcomes as well as determining how changes to the virus genetic material impacts its ability to spread and cause disease.

Data about patients is disparate and diverse, held by many different data controllers. However, the COPI notices put in place due to the pandemic have provided the impetus for a system wide, unified effort to ensure that the data flow channels are established on a nationwide scale. The case studies outlined in Table 2 and Appendix 2 exemplify how genomic research is being propelled by researchers being able to access confidential patient data from every necessary patient, in a timely fashion. The COVID-19 pandemic is likely to be an ongoing challenge and it will therefore be important to facilitate proportionate data sharing in future. As the virus continues to spread and to accumulate genetic changes, it is essential to continue to monitor if these correspond to phenotypic changes in the virus as well as influencing interactions between virus and host genetics.

Whilst it is clear that surveillance and research underpin our efforts to understand the impact, treatment and management of the SARS-CoV-2 virus worldwide, a further question is whether the regulatory arrangements that facilitate these activities are currently ethically and legally acceptable, and whether they will be acceptable in future, once surveillance becomes routine and fewer novel insights can be gained from research.

In the next chapter we turn to consider how patient and public attitudes impact public trust and confidence in creating an effective and proportionate environment for promoting and optimising the generation, collection and sharing of confidential patient data.

4. Patient and public attitudes to the use of confidential data for research

4.1. Introduction

Ensuring that the governance around the collection and use of health data is aligned with the views of patients and publics is vital in order to facilitate transparent and trustworthy data sharing. In this section, we explore patient and public attitudes towards the use of confidential patient information for health research during the pandemic and more generally. We draw on findings from our own focus group, as well as those from other qualitative research studies. There is already a considerable body of evidence on patient and public attitudes to genomic data processing and sharing, which is generally viewed as particularly sensitive by individuals. In this section our focus is the impact of the COVID-19 pandemic on patient/public attitudes. Since restricting our scope to genomic data would have curtailed discussions in our focus group and limited the wider evidence we can draw on, we have broadened our scope to include consideration of views on use of confidential patient information for research more generally. As far as is possible, we aim to analyse how views on common themes such as the need for transparency, trust, data security and public involvement are impacted by the pandemic and consider the implications for data governance going forward. In doing so we aim to identify the considerations that matter to the public, and reflect upon what they think should happen when the pandemic is over.

4.2. Focus group

To supplement our expert interviews and desk based research, PHG Foundation commissioned Traverse Ltd (an independent research organisation) to facilitate a small, two-hour focus group exploring public attitudes towards data sharing during the pandemic. The aims of the group were to identify (i) what considerations were important to the public regarding confidential patient data being collected and used for healthcare and research more freely than usual during the pandemic, and (ii) what they thought should happen when the pandemic was over.

There has been other recent empirical work looking at public attitudes to data sharing, both prior to and during the pandemic: most recently *Data Sharing in a Pandemic: Three Citizens' Juries – Juries Report* exploring public attitudes to pandemic data initiatives. Our focus group was not designed to replicate this significant and demographically representative empirical research but to be complementary as far as is possible and to provide a snapshot of the public's views and primary concerns during this dynamic period, thereby testing for any signs of shifting attitudes.

Traverse facilitated the focus group in June 2021, and their final report is included in full in Appendix 3.

4.2.1. Methods

The focus group took the form of a two hour online session with 10 participants. Traverse used their recruitment partner, Riteangle, to identify and select engaged participants for the focus group. These participants ranged in their age, gender, ethnicity, and socio-economic status, but had all been selected on the basis of having had contact with their healthcare provider in the last 12 months.

Participants were provided with information about changes to information governance at the start of the session, but were not sent materials in advance in order to allow collection of their initial, uninfluenced views. Participants completed a brief pre and post group questionnaire to identify if, and how, their attitudes changed following the additional information and opportunity to discuss their concerns during the session.

4.2.2. Key conclusions

Whilst the focus group involved a small sample, it generated some considerations that have not been widely emphasised elsewhere, although many of their key findings revealed synergies with other similar public attitudes research. The report in Appendix 3 sets out the main topics and discussions in more detail. Key issues and themes included:

Themes from the focus group (extracted from the Traverse workshop report, Appendix 3)

In discussion, the majority of participants were favourable towards the use of patient information for research. They saw the use of patient data in medical research as essential for delivering medical treatment, improving our understanding of health, supporting medical advancement and planning for the future.

However, they did raise some key concerns, and wanted to see specific conditions in place to ensure their data was used appropriately. These surrounded:

- **How their data was kept:** they wanted assurances that their data was kept secure from data breaches, leaks or losses.
- **Who had access to their data:** levels of trust varied across different organisations, with participants expressing widespread trust in the NHS. Trust in the government varied across participants, and trust in commercial organisations was particularly low.
- **Why their data was being used:** participants wanted their patient data to be used for health-related purposes, and to be of public benefit. They were concerned around function creep or change of purpose.
- **What data was being shared:** participants wanted their data to be shared in a non-identifying and

anonymous way, ideally as part of a large dataset. They also wanted to be certain that the data being shared was proportionate to the research being conducted, and that only essential information was shared.

Going forwards, participants identified key conditions to increase their trust in the process of data sharing:

- Participants wanted clear and transparent information about how their data was being used, what ‘anonymisation’ means, and how this might impact them.
- Participants wanted to feel like they had a choice over how their patient data could be used – ideally by consenting for its use in individual medical research projects.

Due to considerable overlap between these key themes and the findings from other studies during the pandemic, in the rest of this section we integrate our findings from the focus group and insights from our interviews with relevant evidence from other studies and reports identified through desk-based research.

4.3. Key considerations affecting public attitudes towards data sharing in a pandemic

The last decade has seen a large number of studies exploring public attitudes to data sharing for healthcare and research. As we discuss in this section, the overall results of these studies show that the public is broadly supportive of data sharing, as long as several conditions are met. First, the primary goal of the research must be to promote the public interest. Second, there must be sufficient safeguards in place to protect the privacy of data subjects and prevent the misuse of data. And finally, there must be trust in the organisations collecting and using data, and transparency around what it is being used for.

This section of the report will consider some of the factors that influence public attitudes towards data sharing, and explore how these have been affected by the current context of the pandemic and the introduction of the COPI notices.

4.3.1. Transparency and communication

Transparency is a core principle underpinning the ethical governance of data¹¹⁶ and is widely recognised in UK healthcare policy and genomics initiatives.

Work done by Understanding Patient Data (UPD) has highlighted the importance of transparency in achieving fairness when it comes to data sharing.¹¹⁷ For transparency to be meaningful, information about the things that people care about should be understandable and accessible¹¹⁸ and therefore one of the ways of delivering this is through effective communication. Whilst there is some publicly available

information online about the COPI notices, what they do, and their impact on the National Data Opt-out, they have not been proactively publicised. With the onus on the public to search for this information, it is likely that only a small proportion will be aware of the notices and their impacts. This concern was echoed within our focus group where participants felt that they should have been informed about the changes brought about by the COPI notices; there should have been more public facing information and engagement around COPI notices and what actions are enabled by them. This sentiment was shared by the citizens' juries, who also expressed unease about the absence of clear information surrounding some pandemic data initiatives and cited 'lack of transparency' as the main reason to oppose them.¹¹⁹ Findings from this empirical work indicate that transparency is important, even during a pandemic and despite the urgency of data collection needed to derive information that could advance understanding of the COVID-19 virus. In fact, it is arguably even more important when public involvement is not feasible in the early stages of change when decisions have to be made quickly.¹²⁰

Transparency requires more than just information provision however, as this neglects the emotional and relational aspects of what makes someone trust, or mistrust, information they are receiving.¹²¹ Distrust of evidence-based advice on vaccinations is a good example: relaying accurate scientific information about vaccine safety and effectiveness is not itself sufficient to address some people's concerns. Interviewees suggested that public engagement is more effective; that the opportunity to ask questions and interrogate leads to increased support for health data use for research (as long as certain essential conditions are met). This was the case in the Traverse focus group, where participants demonstrated slightly higher support for data being shared with the NHS and other public agencies in their post group questionnaire than in the pre group questionnaire. Likewise, there was a small reduction in people who selected that they would prefer their data not be shared at all following the discussion session.

Desire for transparency applies not just to the data sharing initiative in question but includes broader regulatory change, with a desire for information about the purpose of data use, the potential users and secondary uses, in order to understand the personal implications of what is being proposed.

It is possible that concerns about transparency were heightened following the media attention surrounding the planned General Practice Data for Planning and Research (GPDPR) programme, which would allow NHS Digital to collect data on treatments, referrals, and appointments over the past 10 years, alongside other data from medical records data for patients' entire history. This scheme has been criticised¹²² for its failure to communicate with patients in advance of implementation and the backlash it has received may have impacted trust in the COPI notices. It was raised repeatedly by participants of the focus group, perhaps due to the fact that it had considerable media coverage on the day of the discussion. Following this media coverage, it is worth noting that the implementation plans for the GPDPR have been postponed indefinitely.¹²³

4.3.2. Trust and trustworthiness

Decision makers often talk about the importance of building and maintaining public trust in the way that data is collected, used and shared. However, this puts responsibility on the person placing their trust rather than the object of trust. Onora O'Neill argues that the ethical responsibility should lie with the systems and governments. It should not be down to individuals to have responsibility for every instance of their data use. Instead, we should focus on what can be done to make systems, people and institutions worthy of trust through demonstrating trustworthiness.¹²⁴ A number of different features exhibited by 'trustworthy' systems of patient data have been suggested, but characteristics emphasised by Understanding Patient Data include motivation, competence, transparency, governance, accountability and public participation which extend beyond specific instances of data use to wider infrastructures, processes and motivations.¹²⁵ These overlap with some of the principles necessary to satisfy Ada Lovelace Institute's definition of 'participatory data stewardship' which it identifies as the key to trustworthy and responsible data collection and use.¹²⁶ Although the concept of trustworthiness is prominent in academic and policy discourse around the collection and use of data, the participants of the focus group talked in terms of 'trust'. We shall return to the concept of trustworthiness in section 5 of this report.

Trust in institutions plays an important role in the trust participants feel in relation to their research participation.¹²⁷ A key finding from the Traverse focus group is that *who* is using the data seems to be more important than *why* they are using it. Research shows that trust in the NHS is very high, but it is far lower in the commercial organisations.^{128 129} This may be in part due to the fact that people have more faith that the NHS has their best interests at heart or holds values aligned with their own than other organisations. Traverse found that participants were generally comfortable with the NHS using their data for medical research, since they felt the NHS already held this data on them and were highly trustworthy. They were also more supportive of the NHS using their patient data without their direct consent than they were for any other organisation (see Appendix 3).

The public sector, and particularly the NHS, is widely perceived as trustworthy. However in reality the boundaries between public and private sector are not clear cut, and many data driven initiatives require commitment from all relevant stakeholders, including the NHS, government, academic and university researchers, and industry. Increased transparency surrounding these data partnerships and the interplay between these organisations may promote trustworthiness, as "dividing up NHS and 'non-NHS organisations' without reference to purpose can be artificial and misleading."¹³⁰ A full exploration of the trust the public has in each of these stakeholder groups is beyond the scope of this research. In this section we focus on trust in two groups, commercial organisations and the Government, who were the subject of particular attention in our focus group and in other research during the pandemic.

Trust in commercial organisations

It is well established in research on public attitudes towards data sharing in health research that people are uncomfortable with commercial organisations having access to their data, and that this willingness has further declined in the last 5-10 years. A Wellcome Trust survey in 2015 showed that half of UK respondents were willing to share their data with commercial organisations if it was being used for health research purposes,¹³¹ indicating greater levels of support than the results of Imperial College London's survey in late 2018, which found that 95% of UK respondents were not willing to share their medical data with commercial industries.¹³² This rise may be in part due to the increasing strain on citizens' confidence in data sharing strategies following scandals such as Google DeepMind and Facebook-Cambridge Analytica. However overall unwillingness is likely to be influenced by a variety of factors including concerns around a loss of control of their data,¹³³ fears around commercial organisations' motivations,¹³⁴ and how they might use the data once they have access to it.

Some research findings suggest that mistrust is fuelled by a lack of understanding about involvement of commercial organisations, their partnerships with the NHS and their motivations. The type of organisation can trigger patients to make judgements about the purpose of the activity. In short, the 'who' can lead to assumptions about the 'why'. In the 2016 Wellcome Trust survey, just 16% had some awareness (but little depth of understanding) about the involvement of commercial organisations in health research. They also found that if the overall purpose of the data-sharing activity is considered acceptable, concerns relating to the commercial nature of the organisation(s) involved often fade. In the quantitative findings, nearly half (43%) wanted commercial organisations to show 'a clear intent that research will lead to benefits for wider society.'¹³⁵ The blurred lines between traditionally private and public sector ways of collecting data has led to what Ipsos MORI describes as 'context collapse': commercial involvement in health research settings creates a new context that people are unsure how to navigate.¹³⁶ However, research has also shown that the more informed people are, the more likely they are to approve of their health data being used for other purposes, including by commercial organisations.¹³⁷

It appears that concerns about commercial access have persisted despite the pandemic. An example of this can be found in the citizens' juries, where jurors were more supportive of the OpenSAFELY data initiative than others, as it was "developed by doctors, funded by Wellcome Trust grants, and is not currently reliant on commercial funding. The initiative is therefore, by its design, more transparent and accountable as opposed to an initiative created by a commercial third party."¹³⁸ In our focus group, Traverse found that trust in private or profit-driven companies was particularly low. Organisations such as pharmaceutical companies, insurance companies or "American private healthcare firms" were cited as examples of organisations that people would not want to have access to their patient data. Participants worried that these organisations would sell that data on for money or that large corporations would use personal data, as one participant put it, as a "bargaining tool". Participants also did not trust the motivations of these

organisations and doubted that they would work towards the uses of patient data they approved of (medical advancement or planning for the future) (see Appendix 3).

Trust in the UK Government

Trust in the UK government, both in the citizens' juries and the Traverse focus group, was more variable, with very high levels of trust exhibited by some participants and much lower levels from others. This lack of consensus is also significant in the National Data Guardian's Office poll, where a majority (64%) said that they would trust government agencies to use information about them such as coronavirus test results. However, a further 17% did not agree with this and 19% were not sure.¹³⁹

Some initiatives introduced in the pandemic may in fact have eroded trust in the government's management of data. Focus group participants cited difficulties encountered by the NHS Test and Trace app, as well as issues with identifying people on the shielding list for support. The recent media coverage of the GDPR scheme also may have impacted responses; participants said that the fact that there was little notice and little information on this change negatively affected their trust in the scheme, as well as trust in the management of their patient data in general. This is reflected in the uptake of the National Data Opt-out—a service introduced in 2018 allowing patients to withdraw their data from being used for any purpose beyond their immediate care. More people registered for the National Data Opt-out in May 2021 (107,429), when plans for GDPR were released, than in the preceding 10 months (72,225).¹⁴⁰ Following the public furore around the scheme, this figure escalated to nearly 12-fold (1,275,153) in June 2021, taking the number of opt-outs to more than 3 million—almost 5% of the population.¹⁴¹ Bharti et al. point out that this staggering increase in the number of opt-outs shows that despite the legal cover provided by data protection law and the common law duty of confidentiality, NHS Digital did not appear to have secured the 'social license' for GDPR, which is contingent on people's perception of this enterprise being in the public interest.¹⁴² These reservations demonstrate that people still care about what happens to data about them during a pandemic, and that best practice principles of transparency and public involvement still apply.

The role of communities

The role of community or group dynamics is relevant to discussions about trust. The House of Commons' report on *Government transparency and accountability during Covid 19* found that "people are more likely to trust people who they see as "one of us" rather than "one of them"" and the Committee heard that "it is often the behaviour displayed in our communities that influences our own behaviour."¹⁴³ Therefore whilst it is important for there to be a national narrative about how data is used, communications might be best delivered at a local or regional level within a small "diameter of trust"¹⁴⁴, allowing people to ask questions of health professionals and other individuals that they know and trust.

There is an absence of trust in some communities that are disproportionately negatively affected by government policies. Interviewees commented that marginalised groups may be less inclined to deposit their data into a system from which they do not benefit, or trust that the government is acting with their best interests in mind. This could further exacerbate health inequalities: if more people from particular demographics opt-out, the risk of the resulting database being less representative increases.¹⁴⁵

Impact of societal and cultural norms on trust

Societal and cultural norms play a key role in determining baseline levels of trust. In the focus group facilitated by Traverse, participants' views on the acceptability of data sharing (both in general and specifically of patient data) depended on what they had been exposed to, and therefore viewed as 'normal'. For example, participants from countries that routinely collect and share data were more supportive of data sharing during the pandemic than other participants. This is consistent with other studies which have found that European patients' expressions of trust and attitudes to risk were often affected by the regulatory and cultural practices in their home countries.¹⁴⁶

Past experiences of disease and healthcare

Current and past experiences of disease also influenced participant's perspectives on data sharing, with participants who had been diagnosed with health conditions displaying mixed opinions. Some spoke favourably of the potential for data sharing to improve existing treatments and enable the development of new ones, and had benefited first hand from data sharing between different parts of the health system (see Appendix 3). Others however, in particular those with more stigmatised conditions, were more cautious about the potential of patient data sharing, since it held more potential for discrimination.

4.3.3. Purposes and uses

Public benefit

There is strong public support for using patient data to further research and improve care.^{147 148 149} Decades of research have shown that there is great willingness to share data for the purpose of public benefit, although this is not unconditional. The imperative to share data for public benefit seems even more pressing in light of the pandemic. In a press release from the National Data Guardian's Office reporting polling to gauge public opinion on the use of data during the COVID-19 pandemic, 78% of the 2114 people polled agreed that during a public health emergency such as COVID-19, it is more important than usual that health and care data is shared with all those involved in the emergency response.¹⁵⁰

What counts as 'public benefit' can be extremely broad. The 2018 UPD report *Data for Public Benefit* identified five key features that data sharing initiatives designed for public benefit should be able to demonstrate, and concluded that developing a shared understanding of the public benefits people want and expect from data use, and a consistent language with which to talk about them, is vital.¹⁵¹ They also

looked at different types of benefits that individuals value most highly. The number of people able to benefit, the perceived level of need and long term impacts on the services available to people were all seen as important factors. In the Traverse focus group it did not seem to be important whether data was shared specifically for COVID-19 purposes, as specified by the COPI notices, or for health more broadly. What is clear from the combined qualitative research is that public benefit needs to be driving the use of data sharing for research, and where it is motivated by other factors such as financial gain, it is seen as unacceptable.

Function creep

When thinking about data use, participants of the Traverse focus group were also concerned about the possibility of ‘function creep’ or use of data for purposes unintended by the data subject. They wanted to be sure that their personal data would only be used for the purpose(s) they had agreed to, and were concerned that their patient data would be used for non-health uses without their knowledge or consent (see Appendix 3). In practice, this is difficult to guard against, especially during a pandemic. Some boundaries might be set by the scope of any informed consent provided, in terms of the intended use of the data, by whom and for what purposes, and on reconsenting participants as this evolves. But even where there is clarity about the terms of a specific consent, it is not always easy to predict future uses, and data can have unforeseen value either on its own or when combined with other data sets. Indeed in many of the scenarios evaluated by this research, the COPI notices mandate data sharing, thus rendering consent irrelevant, both as a legal basis for data sharing or to ratify the sharing of confidential patient data.

Embedding public views and values in decisions about how NHS data is used could help alleviate concerns about function creep. Ensuring that NHS and patient benefit underpins all partnerships with companies using NHS data will be an important starting point to alleviate these concerns. As raised above, deciding what constitutes ‘public benefit’ is nuanced, and so creating a culture of public involvement in decisions about data and including people from different backgrounds may help to reassure the public that data is not being used in ways that undermine common social goals.

Potential for discrimination

Fear of discrimination, stigmatisation, exploitation or other repercussions as a consequence of data being shared is widely cited in research,^{152 153} with marketing and insurance companies being amongst those that patients are most wary of sharing data with.¹⁵⁴ Evidence suggests that these concerns are still pertinent during the pandemic. For example, a nationally representative survey commissioned by Ada Lovelace Institute found that over half the UK public (54%) think it is likely that vaccine passports would lead to discrimination against marginalised groups. Concerns about bias and discrimination were also a consideration for those who did not use a contact-tracing app.¹⁵⁵

4.3.4. Consent and choice

Consent and choice are vital to discussions about data sharing, as whilst many people are supportive of using patient data to improve health research, they often think that they should be asked to give permission first – either through an opt-in or consent based system.¹⁵⁶ Previous research in public dialogues show that patients would prefer to be asked to consent to each instance of data use.¹⁵⁷ This was reflected to some extent in the Traverse focus group where participants voiced their preference for control over how their patient data is used through consent mechanisms, even during the pandemic (see Appendix 3). Undoubtedly informed consent is the right model in some instances, such as participation in clinical trials, where active and ongoing participation is required and the risks associated with involvement may be high. However, relying on informed consent for all uses of data from patient records is unlikely to be feasible and would place a disproportionately high burden on patients, and research shows that when patients understand what this means in practice – frequent contact and engagement – their views can change.¹⁵⁸

Instead, they want to know that individuals and organisations making decisions about the use of their data are trustworthy, and do so with their interests in mind. The National Data Guardian Review of Data Security, Consent and Opt-Outs noted that *'Most people do not feel the need to know what is happening with their data, and people want to be able to trust the system and know that everything is okay'*.¹⁵⁹ This is another example of the importance of trustworthiness as a precursor to trust. Ethical responsibility should lie with systems and governments, rather than requiring individuals to take responsibility for and consent to every instance of their data use. In the UK, independent review committees such as the Confidentiality Advisory Committee (CAG), research ethics committees and local Caldicott Guardians take on some of this ethical responsibility, through assessing benefits and risks and ensuring a strong case for the public and social benefits of using the data.

The National Data Opt-out

The National Data Opt-out is an important part of the landscape surrounding consent and patient choice. Described in greater depth in section 2, this scheme, introduced in May 2018, allows patients the right to opt-out of their confidential patient information being used for purposes beyond their individual care, thus providing an element of choice to publics and a solution for those who disagree with how their data is being used. However, this opt-out does not apply where confidential patient information is being used to protect public health, as it is under the COPI notices. Therefore confidential patient information can still be used for COVID-19 related purposes, regardless of whether the individual has registered an opt-out.

Consent during the pandemic

The disapplication of the opt-out was easier to justify at the start of the COVID-19 pandemic, as a proportionate reaction to a global emergency, but the COPI notices have been extended three times since

then (as of October 2021). As vaccination reduces the threat from COVID-19, further extension is likely to place greater stress on the original rationale for the COPI notices and their restriction of the opt-out.

4.3.5. Privacy and data security

The use of depersonalised data presents a communications challenge for maintaining trust in data-sharing. The public tend to be happier with sharing data when it is anonymous and cannot be traced to individual patients,¹⁶⁰ but in practice this is hard to guarantee. Whilst the public recognise that data is never completely safe,¹⁶¹ more needs to be done to put measures in place that minimise risk and promote trustworthiness.

Safeguards

Unsurprisingly, participants in the Traverse focus group are most comfortable with personal data being used when it is shared in anonymous, aggregated forms, assessing this as posing a minimal risk to individual privacy (see Appendix 3). Although the participants raised anonymisation as a key requirement for data sharing, they were confused about terms like anonymisation and pseudonymisation and how they differed from each other and from identifiable data. This is exacerbated by the fact that these terms are often used interchangeably, even by researchers, and so disentangling them and communicating them to patients will be a challenge.

In terms of accessing data securely, interviewees proposed that data should be held in a Trusted Research Environment (TRE) in order to minimise data security risks. TREs are secure spaces for researchers to access sensitive data. Rather than receiving downloads of data for researchers to analyse on their own systems, in a TRE, the users go to the data rather than vice versa. This provides greater assurance that the data is handled securely by trusted parties, as data can be tracked and technical safeguards ensure no data leaves the secure environment.¹⁶² However, re-identification risk amplifies with increased data linkage across a wide range of disparate datasets. Interviewees raised the possibility that linking datasets could create a richer multidimensional dataset than may be expected.

It is notable that these safeguards, although important, are not sufficient for trustworthiness, and that people still care what happens to data about their health even if it is anonymised and stored in a TRE.

4.3.6. Public and patient involvement

Increasingly, there is an expectation that there is some element of public involvement in decision making so that public views and values can be embedded in these processes. 'Involvement' is a broad term encompassing different mechanisms for achieving different types and levels of involvement. These are often not mutually exclusive, but include focus groups, surveys, lived experience panels, public deliberation, citizens' juries, co-design and consultation activities.

The benefits of participatory approaches to research are well established. Although public involvement is not a legal requirement for research regulated by the HRA in the UK, it is good practice. The UK Policy Framework for Health and Social Care Research asserts that minimum good practice for health and social care research in the UK is for patients, service users and the public to be involved in the design, management, conduct and dissemination of research.¹⁶³ However the need for some form of public scrutiny and involvement extends beyond individual research projects and initiatives, to data governance and regulatory decision making.

Most importantly, people *want* their views to be taken into account. Despite all the pseudonymisation and anonymisation measures in place, research suggests that publics deeply care about how their data is used: 74% of people believe the public should be involved in decisions about how NHS data is used.¹⁶⁴ People want to know that their views are represented in what can be difficult value judgements about who should get access to their data and why.

With this in mind, new models are being proposed, such as data stewardship which is explored in depth by Ada Lovelace Institute in their 2021 report *Participatory data stewardship*. This outlines the benefits of effective participation in the design, development and use of data and data-governance frameworks.¹⁶⁵ Evidence suggests that public deliberation, when undertaken effectively (early on and with intent to respond from the commissioners of the process) can have a tangible impact in shaping policy outcomes so that they take greater account of public values.¹⁶⁶

In the genomics context, the Chief Medical Officer's 2016 report 'Generation Genome' called for a reframing of the implicit social contract for medical research and medical practice between the NHS and patients/publics.¹⁶⁷ This novel social contract provides a precedent for building upon notions of solidarity and reciprocity, with publics participating and relying upon the trustworthiness of the health system, and the health system earning this trust by accepting responsibilities for improving information security and governance, including transparency and accountability.

Public involvement can be facilitated at national and local levels. Whilst it is important that there is a coherent national narrative about how data is used, it may be easier to meaningfully engage the public and embed community values on a local level than a centralised one. Delivering communications tailored to the local context and allowing people to ask questions of local politicians, experts and healthcare professionals in their community may also help to promote genuine dialogic engagement. This must be distinguished from more narrowly focused public relations exercises that seek to 'capture' the public i.e. to persuade the public of the legitimacy of decisions already taken by experts.¹⁶⁸ Engagement initiatives at a local level are already being implemented and could be built upon, for example One London, or Local Health and Care Record Exemplars.

4.4. The future of the COPI notices

What are the implications of the evidence from public attitudes research, both before and during the pandemic, for the COPI notices going forward?

The results of the post discussion questionnaire showed that the participants of the Traverse focus group did not feel comfortable with the COPI notices being extended permanently, irrespective of whether data was used specifically for COVID-19 related purposes or broader health purposes. Participants feared that “the new guidelines would be kept and ‘quietly forgotten about’” posing risks to their data privacy and freedom of choice. They preferred instead that information governance revert back to its pre-Covid state, but that the data collected so far be kept rather than deleted (see Appendix 3). Interestingly, this was not mirrored in the citizens' juries, where a majority of participants were in favour of all the data sharing initiatives continuing for as long as they are valuable.¹⁶⁹ This applies to the initiatives themselves however, rather than the regulatory changes underpinning them.

It is impossible to draw definitive conclusions about whether or not to extend the COPI notices permanently based on findings derived from a small number of participants. However, exploring which considerations are important to the public, particularly in the current context of the pandemic, can provide tentative insights into what is needed from a regulatory framework.

For example, it is clear in our (and other similar) qualitative research that the public want some kind of engagement or involvement in decisions about adjustments to the regulatory framework. A mechanism for doing that was not suggested in the focus group but in the juries there was some appetite for an independent advisory group comprising of experts and laypeople.¹⁷⁰ Embedding more thorough public deliberation thoughtfully and effectively requires time and resourcing (often over months rather than weeks), and this can be in tension with the imperative to work rapidly in developing and designing data-driven systems, or in developing policy and regulation to govern the use of these systems. The Covid-19 pandemic is an example of where urgent decision-making demands a more iterative and agile approach to assembling data infrastructures. In this instance, deliberative exercises may not always appear to be expedient or proportionate, but might on balance be valuable and save energy and effort if regulatory change is likely to generate significant societal concern.¹⁷¹

When asked about what they anticipated would happen to the COPI notices in the future, some interviewees cautioned against embedding the powers enabled by the notices into the regulatory framework permanently, without first undergoing a substantive process of consultation, engagement and transparency. They argued that it was a mistake to translate what is possible under the COPI notices into ‘this is how we do things now’ as there is no social licence for the use of patient data to be used in this way beyond the pandemic.

4.5. Section summary

The pandemic underscores the need for healthcare data sharing at scale. Data has proven to be indispensable for understanding disease epidemiology, developing effective vaccines and treatments and proactively prioritising sub-populations for support, surveillance or vaccination. Research shows that patients and the public broadly support the use of confidential patient data for health research, especially during a pandemic, but that this support is not unconditional, even during a public health emergency. The response to GDPR has highlighted that the public do still care what happens to data concerning them, and that the same concerns about transparency, data security, accountability and public involvement are still pertinent to their decisions.

Our research has highlighted that, at an individual level, attitudes about data sharing are heavily influenced by people's backgrounds, communities and experiences. It has also reinforced existing evidence that the nature of the individuals and organisations that have access to the data is important. This is due largely to the fact that some are deemed more trustworthy than others, and are accompanied by preconceptions about motivations and what that data will consequently be used for. Dwindling public trust can hamper meaningful adoption of data, and therefore building trustworthiness must be prioritised.

Our analysis has shown that there are a number of interrelated features that patients and publics view as essential for ethical data sharing. Appropriate safeguards are needed to ensure that sensitive data is kept secure. Transparency is key – not just about methods and procedures, but also about expected benefits, risks, and underlying values – and must extend beyond communication to include meaningful public engagement. Beyond legal reassurances, publics need to be involved in discussions about data governance and use – e.g. what amounts to public benefit, the trade-offs they are prepared to make – and then offered means to exercise their choice (for example through providing consent or an opt-out system). If consent is not being used as a mechanism to enable choice, then extra effort is needed to fulfil ethical responsibilities towards people's data.

These features are all interconnected, and inadequacies in one of these areas make others even more essential. For example, choice is seen as an important means of enabling publics to exercise their autonomy and retain an element of control over who has access to their data and for what purposes. However, this becomes particularly important in the absence of transparency and public involvement.

Together, these features promote an ecosystem of trust, which can accelerate the mobilisation of health data for medical research and innovation.

5. Discussion

In this penultimate section of the report we combine our reviews of the legal framework and landscape of data used for genomic and medical research during COVID-19, with our analysis of patient and public attitudes, insights from stakeholder interviews and further ethical and legal analysis to draw conclusions about the impact of the COPI notices and potential changes that may be taken forward.

Research questions:

1. How have regulatory changes to information governance to support research into COVID-19 impacted genomic and medical research?
2. Should part, or all, of the changes be permanently integrated into the regulatory framework?

5.1 How have regulatory changes to information governance to support research into COVID-19 impacted genomic and medical research?

Beginning with the first part of our research question, our review of the legal framework clearly identifies the COPI notices as the central regulatory change to information governance during the pandemic. The COPI notices set aside the common law duty of confidentiality for COVID-19 purposes without consent or any further authorisation being required for sharing confidential patient information (CPI) within the terms set out in the notices. The statutory context for the notices places certain limits around them including:

- an overarching goal of improving patient care or that the processing is in the public interest;
- processing should be for public health purposes, including research into communicable diseases and other risks to public health;
- a range of limitations and safeguards apply including that:
 - information should be de-identified as far as possible and as minimal information as possible should be processed to achieve the permitted purposes;
 - access should be limited to those who are required for and aware of the purposes of processing;

- technical and organisational measures should be taken to prevent unauthorised processing, and;
 - the need to process CPI should be reviewed by the person in possession of that information at least every 12 months.
- A final and important restriction is that the normal requirements of data protection law shall apply without modification.

However, these are broad outer limits and within them, the COPI notices enable a range of processing. Because the COPI notices apply to multiple organisations, there is no single decision-maker or overseer and it is not possible to define the scope of 'COVID-19 purposes'. This is something that may contribute to a sense of a lack of transparency which our focus group identified and which is noted in wider research.

For researchers, the COPI notices provide a new pathway which is separate to explicit consent or the usual pathway for approval without consent under Regulation 5 of the COPI Regulations: a recommendation from the Health Research Authority's (HRA) Confidentiality Advisory group (CAG) and approval by the HRA or, for non-research purposes, the Secretary of State for Health and Social Care.

The impact of the COPI notices on genomic and medical research

Our review of the landscape of data used for genomic research during COVID-19 identifies a remarkable development of data initiatives addressing COVID-19 that have been (in part) reliant on the COPI notices. Many of these have leveraged existing projects, infrastructures and organisations to address COVID-19. They include the COVID-19 Genomics UK (COG-UK) consortium and initiatives established by Public Health England for genomic surveillance, as well as a number of large-scale genomic research initiatives, such as the GenOMICC study, with linkages to Genomics England and COG-UK, the HOCl study initiated by COG-UK, the SIREN study and research enabled by UK Biobank relating to COVID-19. These examples are non-exhaustive and there are likely to be many more data linkages based on the flows of data enabled by the COPI notices.

While our research cannot categorically determine these initiatives would have been limited in their scale, nature, power or speed without the COPI notices, there is evidence that the COPI notices have been essential enablers for pandemic data sharing and our interviews indicate considerable benefits in terms of speed and ease of data access under the COPI notices.

Interviewees emphasised that it could be a challenge to get access to data, including confidential patient data before the pandemic, even for researchers with the participants' consent to access it. They reported hesitancy on the part of data custodians to release data in a timely manner and a complex legal and governance landscape around access. By contrast, interviewees (including some researchers) emphasised the role of the COPI notices in enabling widespread access to data during the pandemic. For several of our interviewees, a major benefit of the COPI notices was felt to be the mandating of data sharing between

certain actors for COVID-19 purposes because it had been challenging to obtain data on a discretionary basis, despite what the researchers felt was a valid authorisation such as the participant's consent, being in place. For some types of research the COPI notices did not necessarily enable access to data that researchers were not already authorised to obtain on the basis of consent, but the interviewees emphasised that COPI notices have made it quicker and easier to access data and that they provide confidence to those sharing that they are acting appropriately. Researchers we interviewed were concerned about the impact on their access to data if the COPI notices expire and are not replaced.

We are not in a position to fully untangle the impact of the COPI notices as a new legal authorisation for research using CPI or as a mandatory requirement for data-sharing between certain actors, from the impact of associated new or streamlined processes and pathways for data-sharing introduced to facilitate COVID-19 data processing. However, the COPI notices can also be seen as a signal across the health data ecosystem that data should be shared for COVID-19 purposes and this in itself may have an important impact on access to data for research.

What we can identify is a considerable appetite for the speed and efficiency of data access enabled during the pandemic to continue into the future. This is likely to give rise to calls for COPI notices to be either extended (in current or modified form) or for other changes to be made which maintain the improvements to data access for research seen during the pandemic.

Consideration: Continued viral and disease monitoring will be needed as the pandemic progresses but more clarity is needed about the distinction between research and surveillance in order to achieve a proportionate regulatory approach.

Answering the question of whether such changes should be permanently integrated into the UK's regulatory framework inevitably requires further work to try and disentangle which aspects of the COPI notices, or the procedures and processes surrounding data-sharing authorised by them, have had the greatest benefits. However, our research indicates a variety of different policy options.

5.2 Should part, or all, of the changes be permanently integrated into the UK's regulatory framework?

The key question is whether some or all the changes brought about by the COPI notices should be taken forwards as part of the long-term regulatory framework. As noted already, a number of different changes

could be proposed with an initial distinction made between a long-term extension of the COPI notices for COVID-19 purposes and other changes.

5.2.1 Longer-term extension of the COPI notices for COVID-19 purposes?

As we discussed in section 3, COVID-19 is not likely to disappear suddenly. Most scientists consider that COVID-19 will become endemic¹⁷² and it is likely that ongoing surveillance and investigation of new viral mutations and their impact on transmissibility, severity or immune evasion will be required in the longer term. Wider research into the long-term health impacts of COVID-19, its interaction with other health conditions and implications for management and care is also likely to be highly important and continue for the long-term.

The NHS Health Research Authority has issued guidance for the transition of research from COPI notices to the well-established Regulation 5 support authorisation process (see section 3),¹⁷³ which may imply an intention for all COVID-19 research to cease reliance on the COPI notices at the end of March 2022. It is currently unclear if this will apply to all forms of research or whether some longer-term COVID-19 research (or indeed other forms of research as we discuss below) may seek an extended form of COPI notice, perhaps modified to mandate specific data flows between specific actors.

Key considerations

Although some justification for regulatory exceptions for COVID-19 purposes may continue in the short to medium term, this is not likely to be the same justification as was originally the case early in the pandemic or at the renewal of the notices, as vaccination and behavioural measures are significantly reducing the threat to public health. For some, including a number of our interviewees, the COPI notices should be seen as an emergency measure which should only continue as long as COVID-19 continues to be a public health emergency. This was echoed recently by the National Data Guardian, Dr Nicola Byrne, who stated: ‘We need to be very clear that the emergency use of emergency powers is short term and only to the end of the pandemic, and for the purposes of the Covid response. There has to be an endpoint.’¹⁷⁴

It is important to distinguish between the justification for the COPI notices on the basis of the threat from COVID-19 and other potential justifications for retaining or extending elements of these regulatory changes, for example to address the high burden of other diseases such as cancers and cardiovascular disease. It could be argued that there are equally compelling alternative justifications for changes to data access for health research beyond COVID-19 but the crucial factor highlighted in section 4 is patient/public desire for transparency and engagement about such justifications and the processing they enable.

The clear message from our focus group and wider work on patient/public attitudes to data sharing in the pandemic is that there is widespread support for data sharing to tackle COVID-19. However, patients/publics also expressed a clear desire for greater transparency and engagement about the

processing that is enabled by the COPI notices. We will turn to the issues of trustworthiness, transparency and engagement more fully below.

Consideration: If it is determined that continued use of the COPI notices – or similar alternative measures enabling processing for COVID-19 purposes- are proportionate, there should be increased transparency and public engagement around the uses of patient data for COVID-19 purposes.

Allied to the public desire for transparency of processing enabled by the COPI notices is the fact that the scope of ‘COVID-19’ purposes is potentially very broad and it would be beneficial to have further clarity about processing that is within or beyond the scope of such purposes. This may require that an accountable body is designated to have oversight of these purposes and to make further decisions about processing that is within the scope of the COPI notices.

5.2.2 Wider reform of the regulatory framework for genomic and health research using confidential patient information?

The COVID-19 pandemic has been an unparalleled emergency and there is a strong justification and public support for measures taken to enable data processing addressing COVID-19 and its impact on the health system. Not only researchers but also the public have become aware of the power of health data access for vaccine development, variant tracking and other research with far-reaching benefits for millions of people.

The goal for regulators and policymakers should be to build on this experience with measures that support genomic and health research more widely, while maintaining public and professional support and confidence. A synthesis of our legal analysis, scientific review, focus group, interviews and analysis of public attitudes research suggest a number of key considerations that should be taken into account in determining whether, and how, changes to the regulatory framework should be made. Before exploring these considerations further, we first outline some possible changes and the specific issues that they may generate. These proposals are not exhaustive and it is beyond the scope of our research to evaluate the strengths and limitations of hypothetical changes in detail.

What reforms or changes may be proposed?

Based on the current COPI notices, a range of potential changes to the regulation of confidential patient information for research purposes could be envisaged.

A new legal basis for research using CPI without consent

The COPI notices currently provide an exception to consent for COVID-19 purposes. However, as discussed in section 2, Regulation 5 of the COPI Regulations also provides a limited exception for research purposes (and for some non-research purposes) on application to the NHS Health Research Authority's Confidentiality Advisory Group (CAG). In fact, our interviewees highlighted that CAG has had an active role in continuing to scrutinise research ethics applications during the pandemic where they have raised important issues of confidentiality. None of our interviewees called for the removal of this form of scrutiny for research. It may be that CAG could continue to build on the experience of expedited review to further streamline the process for certain cases, but it is not clear that any stakeholders would favour an alternative legal mechanism that removes this oversight.

Mandatory data sharing for specific purposes or between specific actors

The more significant change implemented by the COPI notices has been the removal of discretion and the mandating of data sharing for COVID-19 purposes when requested by specific parties. The most significant example of this highlighted by our interviewees has been the mandatory sharing of primary care data. This was contrasted with the pre-pandemic situation by several interviewees who described considerable difficulty obtaining primary care data. This was reportedly due to the difficulty persuading data custodians of the validity of a consent that had not been obtained by them directly and their reluctance to accept the assurance of the researchers as sufficient to authorise disclosure of confidential patient information.

As some interviewees noted, this highlights a distinction between the individual nature of the duty of confidentiality, which applies at the level of the individual professional, and responsibility for data protection, which applies at an organisational level. This means that professionals may be understandably reluctant to disclose CPI without considerable assurances because if disclosed they will be individually liable for a breach of confidence if the consent or other authorisation is found to be invalid.

Consideration: As the duty of confidentiality applies at an individual- as opposed to organisational-level, mandating the disclosure of confidential patient information is only likely to be accepted by healthcare professionals if it is accompanied by considerable transparency and assurance of the validity of the authorisation, its scope and the safeguards that apply.

There are multiple possibilities for more limited mandating of data sharing. These include mandating sharing for certain purposes, as has been the case with COVID-19 purposes in the COPI notices. In this research we have a particular interest in the impact of the COPI notices on genomic research but none of our interviewees suggested that there is a special case for requiring that CPI is shared for genomic research

purposes, as opposed to any other research purpose. Another possibility is mandated sharing but only with very specific recipients, for example, Trusted Research Environments (TREs). The controls and technical safeguards provided by TREs are designed to significantly reduce threats to privacy and confidentiality while enabling research. The citizen's juries exploring pandemic data initiatives (discussed in section 4) identified considerable approval and support for the approach taken by OpenSAFELY facilitating research using GP data in a TRE during the pandemic.¹⁷⁵ It is notable that the plans for the General Practice Data for Planning and Research initiative have now committed to only allowing access to data via a TRE.¹⁷⁶

However, regardless of the nature of proposed changes to regulation of CPI for genomic or health research, our research has identified a range of ethical and legal considerations which should be taken into account.

5.2.3 Key ethical and legal considerations for changes to regulation of CPI for health research

In section 4 we discussed factors that influence public attitudes towards data sharing in the pandemic. Combined with findings from the legal analysis, scientific review and insights from our interviews, these factors lead to a set of key ethical and legal considerations which are relevant to any extension or change to the regulation of CPI for research purposes.

Trustworthiness and fostering confidence in data sharing

As the GDPR initiative and the care.data experience before that demonstrate, a failure of patient and public confidence in proposed reforms can critically delay or completely terminate plans for improvements to data access, linkage and sharing for important health purposes. We agree with our interviewees who adopted Onora O'Neill's argument, that it is less appropriate to talk about building and maintaining trust than it is to consider how to make systems, people and institutions worthy of trust. There is generally higher trust in the NHS and doctors than there is in either the government or commercial organisations and the challenge for demonstrating trustworthiness in data reforms at a national scale, is that all these parties are likely to be involved.

Consideration: Investing time and resources into promoting characteristics of trustworthiness – such as transparency and public engagement – may help generate collaborative agreement on how data should be used and for what purposes.

Transparency, engagement and involvement

A key starting point emphasised by our focus group and in wider empirical research is transparency. Even during the pandemic with widespread support for measures to address COVID-19, participants desired more information about what was being done with their data under COPI notices. As our interviewees also emphasised, information should also be provided about the opportunity costs of not linking or sharing data. However, what is required is more than just information provision. A clear finding from the combined research (section 4) is a strong desire for public engagement and involvement in decisions about data.

Consideration: Good public engagement must move beyond informing people, and towards enabling patients and publics to shape decisions about who has access to data concerning them, and for what purposes.

Facilitating engagement and involvement for national level decisions may not be easy, but a number of approaches have been suggested. Understanding Patient Data highlight examples of good practice using online deliberation, a citizen's summit and citizen's juries, and emphasise that a range of methods can be used within one overall coherent approach.¹⁷⁷ There is no single public but instead a range of publics with different perspectives and experiences, including those with less trust in government and decision-makers. As we discussed in section 4, this may require engagement at a local level, across communities, to address the complexity and variety of public concerns.

There can be a tension between the need to develop new approaches to data as quickly as possible, for example in the pandemic, and embedding thorough public engagement and deliberation. As research has found, there is greater acceptance for measures taken in this context. However, for longer-term and population wide proposals, such engagement may be essential if regulatory change is to be accepted by patients, professionals and the public.

Consideration: Widespread efforts should be made at national and local levels to meaningfully engage publics and different communities in decisions about changes to the regulation of confidential patient information.

The role of consent and choices about data

Consent and choice in relation to use of confidential patient data is not straightforward. Requiring active, opt-in consents can become very burdensome for individuals and it places ethical responsibility squarely on individuals rather than with the systems and decision-makers who should be taking steps to demonstrate their trustworthiness to the public. The National Data Opt-out was felt to strike the right balance by setting a default that most people are happy for such use of their data but also allowing a choice to be expressed if they do object. However, as means of giving individuals a choice over the processing of their CPI, the National Data Opt-out is a blunt instrument. For example, it does not allow individuals to opt out of processing by certain types of actors, such as private sector organisations, or to make more granular decisions about sharing, or opt-out for part of their CPI as opposed to all their health information. It also does not apply to disclosure of anonymised information. On one hand this reduces the potentially complex impact of different choices on datasets and analysis. However, it may also force a binary decision to withhold sharing despite individuals only being concerned about a specific aspect of their data or its potential use.

One possibility is to provide for more specific choices about data use. However communicating more granular choices is difficult, as is building systems to facilitate and maintain downstream commitments. Another possibility is to identify and respond to specific concerns through engagement and deliberation. For example, the role of commercial parties in health research and development is highly complex and would benefit from more nuanced conversations with publics.

In the context of consent to disclosure of CPI, we have identified that this can be highly problematic even when a specific and informed consent has been obtained because the validity and scope of that consent has to be persuasively communicated to (potentially myriad) data custodians. This is an area that could benefit from guidance about what constitutes acceptable and valid consent to disclosure of confidential patient information.

Consideration: In order to operate effectively as a mechanism to authorise disclosure of CPI for genomic and medical research, there needs to be a shared consensus between healthcare professionals (including GP's) and researchers about what constitutes a valid consent and the evidence required to reassure other professionals of that validity.

Privacy, security and data protection

Securing privacy and safeguarding data is paramount. Our research confirmed findings from wider empirical evidence that people are most comfortable with data being used when it is in anonymous form. However, as we discuss in our legal analysis, defining what is meant by anonymous is complicated and potentially differs depending on whether data protection law or the law of confidentiality is being considered. Most uses of data for genomic and health research utilise pseudonymous data, where identifiers have been removed and separated from data. Such data remain identifiable to those who hold the 'key' so, although highly safeguarded, there are still risks of identification. It was clear in our focus group that there is confusion about these different terms and states of data.

Being open and transparent about these terms and that there will almost always remain a residual risk to privacy may be important in managing the public's expectations and for building trustworthiness and confidence. Otherwise, there is a danger that confidence will be lost if data which the public assume to be anonymous are shown to be identifying.

Technical safeguards

A related consideration is that regulatory reform may not be required if there are technical measures and safeguards that can be used to achieve the same goals. For example, it may be that technical encryption measures and de-identification methods can be used to avoid the need for linking or sharing confidential patient information for some research purposes entirely. As discussed earlier, Trusted Research Environments hold great promise in this regard but there should be caution in assuming that data held in these environments are no longer CPI particularly when this could impact public trust and confidence.

The interaction between different regulatory domains

One challenge identified in our scientific review, legal analysis and in our interviews is the complex interaction between different domains of regulation and related activities. For example, there are blurred boundaries between some forms of research and activities that may be considered public health surveillance. This can be challenging for researchers and others to navigate. Further guidance on the distinction between research and public health activities is needed from bodies such as the HRA and those parts of government that have taken on Public Health England's responsibilities overseeing surveillance.

As well as a cause of confusion, the boundary between public health and research may be important if changes are proposed which lead to perceptions that one route to obtaining data is harder than another.

Consideration: Care should be taken that proposals for data governance in one domain, such as in public health or in research, do not incentivise actors to use one route over another inappropriately.

Another major area of confusion relates to the interaction between data protection law and the common law of confidentiality. As our interviews found, this can be complex and uncertain for researchers and professionals, let alone for the public who may be informed that their consent is not necessary for data protection purposes, for example, but that it is required for disclosure of CPI. Moreover, there is debate between experts in these areas about the appropriate interpretation of key concepts such as ‘personal data’ and the scope of CPI which means that determining how they align is currently unclear. This is even more uncertain following Brexit and the dynamic situation for UK data protection law, with a live consultation on its potential reform.¹⁷⁸

As some interviewees discussed, it may be that ‘personal data’ as conceptualised by data protection law has evolved with the introduction of the UK GDPR to incorporate a wider range of potentially identifiable data than previously thought (see section 2). But the statutory definition of CPI may continue to be interpreted as it was prior to the GDPR, leading to considerable divergence, in particular with CPI not considered to extend to pseudonymised data in the hands of recipients who have no access to a code or ‘key’. However, there is also the possibility that the courts could interpret the scope of CPI more restrictively if called on to adjudicate.

In many ways, the approach taken to the scope of CPI is in the hands of NHS organisations in terms of the standards and guidelines they apply. It may not be feasible or appropriate to align this with data protection law but, as already discussed, transparency about differences is important for both professionals and publics.

Consideration: The development of guidance and transparent information about the interaction between data protection law and the common law of confidentiality in the research context could help professionals and the public understand and navigate this complex terrain.

6. Conclusions

This report considers how the regulatory changes to the governance of confidential patient information for COVID-19 purposes—the COPI notices have impacted genomic and medical research, and whether these changes should be integrated into the regulatory framework. Since the start of the pandemic less than two years ago there has been a phenomenal growth of initiatives, research programmes and public health activities that harness CPI for COVID-19 research and surveillance. Many of these have relied on the COPI notices to authorise the disclosure and linkage of confidential patient information. Insights from our interviews with stakeholders at the research and data coalface suggest that the COPI notices have been essential enablers for pandemic data sharing, greatly increasing the speed of data access and linkage and, by removing discretion from data custodians, unlocking some CPI that had been difficult to obtain prior to the pandemic. However, the impact of the COPI notices is likely to be a combination of the regulatory changes they have made, the new or streamlined processes that have been developed to facilitate COPI notice-authorised data access and the powerful signal they have sent about the importance of data sharing and access to combat COVID-19. This means that we cannot draw firm conclusions about whether the regulatory changes the COPI notices have made should be made permanent.

We can, on the basis of further ethical and legal analysis and consideration of public attitudes from our own focus group and wider empirical evidence, draw conclusions about the key considerations that should be taken into account in any decisions about extension of the existing notices or any wider reforms based on them. In terms of extending the COPI notices beyond March 2022 there should be increased transparency and public engagement around the COPI notices and the processing they enable. At present the COPI notices enable a potentially broad and unclear range of processing for COVID-19 purposes. This may benefit from the oversight of an independent body if the notices are to be extended long-term. A failure to address this risks harming public confidence in the notices and the data ecosystem more generally.

Drawing on the experience of the COPI notices, a potentially wide range of reforms to the regulation of CPI for research purposes could be envisaged. These include providing a new legal basis for research using CPI without consent and mandating data sharing between specific actors for specific purposes. Our research emphasises that, irrespective of the proposed reform, a number of key ethical and legal considerations should be taken into account. These revolve around the central importance of developing and demonstrating trustworthiness in data sharing and include the importance of transparency, public engagement and involvement in both specific and general decisions about what is done with confidential patient information. Choice and consent are also relevant but it is primarily the responsibility of data custodians and decision makers to create the conditions for public trust and confidence in the use of patient data. Safeguarding privacy is also crucial and the development of Trusted Research Environments is very positive from this regard.

One consideration that arises in particular from our legal research and interviews with stakeholders, is the need for an assessment of the impact of changes to the regulation of patient data in the complicated context of overlapping laws and legal requirements. Two areas of overlap of particular importance are between data protection law and the common law of confidentiality, and between research and public health surveillance activities.

A way ahead?

Ultimately, these considerations are heavily interconnected and collectively will play an important role in achieving the objectives set out by Government in the National Data Strategy¹⁷⁹ of maintaining the ‘high watermark of data use set during the pandemic’ and harnessing the power of data for improvements in health and to drive innovation and growth. The draft data strategy for health and care produced by NHSX emphasises the transformational power of data for health and care,¹⁸⁰ and sets out a range of commitments and proposals which take into account the considerations outlined above. In particular, there are ambitions to increasingly look to use Trusted Research Environments for secure health data analysis and to produce guidance to address the complexity and uncertainty inherent in the information governance landscape. There are also proposals to bring people closer to ‘their data’, including enabling everyone to access the health and care data which is about them.

While these general ambitions are to be welcomed, it is clear from our research that more than providing information about what is being done with data is required to realise these strategic data objectives. With recent examples in mind, it is clear that a more profound discourse around the use of patient information and a concerted effort at widespread engagement will be necessary to foster professional and public confidence in the actors, processes and infrastructure of which the system is composed.

This echoes the discussion in the genomics field about the need for a broad rethinking of the ‘social contract’ for medical practice and research in the UK in order to secure the benefits of genomic medicine.¹⁸¹ As Jonathan Montgomery, Anneke Lucassen and Michael Parker set out in their Chapter of the Chief Medical Officer’s 2016 Annual Report, *Generation Genome*, ‘[t]o achieve this, processes for creating common understanding are required, as well as mechanisms for revising the agreement when necessary.’

Finally, a range of alternative legal mechanisms for data stewardship are now being discussed in the health data space and more widely. These include data trusts, data intermediaries, data cooperatives, and alternative contractual or corporate models.¹⁸² These promise to provide individuals with more control over how their data are used and the terms on which confidential information may be disclosed. These mechanisms could have an important role in enabling trustworthy access to data for research, perhaps for specific communities or in relation to particularly sensitive areas of research. The central place of the NHS in the UK’s health and social care system provides unique opportunities for the UK to harness an unprecedented amount and quality of routinely collected health data for research. Maximising the

potential of those data on a population scale requires proportionate safeguards coupled with unequivocal national level commitments to transparency and public dialogue.

References

- 1 General Medical Council. Confidentiality: Good practice in handling patient information. 25 April 2017 [Last updated 25 May 2018].
- 2 The Caldicott Committee. Report on the Review of Patient-Identifiable Information. [online] Department of Health. 1997.
- 3 Caldicott F. Information – to share or not to share. The Information Governance Review. Department of Health. March 2013.
- 4 National Data Guardian for Health and Care. Review of Data Security, Consent and Opt-outs. June 2016.
- 5 Gov.UK. UK Caldicott Guardian Council. Available from: <https://www.gov.uk/government/groups/uk-caldicott-guardian-council>
- 6 National Data Guardian. Guidance: The Caldicott Principles. December 2020.
- 7 Topol E. The Topol Review: Preparing the healthcare workforce to deliver the digital future. An independent report on behalf of the Secretary of State for Health and Social Care. 2019.
- 8 Thomas R, Walport M. Data Sharing Review Report. 11 July 2008; Department for Business, Energy & Industrial Strategy. Industrial Strategy: Building a Britain Fit for the Future. November 2017.
- 9 SNOMED CT is a structured clinical vocabulary for use in an electronic health record. [NHS Digital. SNOMED CT. [online] January 2021. Available from: <https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct>]
- 10 Vezyridis P, Timmons S. Understanding the care.data conundrum: new information flows for economic growth. *Big Data & Society*. 2017; 1-12
- 11 Greenhalgh T et al. Adoption and non-adoption of a shared electronic summary record in England: a mixed-method case study. *BMJ*. 2010; 340:c3111.
- 12 Warren-Gash C. Linking and sharing routine health data for research in England. PHG Foundation. 2017.
- 13 These juries were designed by Dr Malcolm Oswald of Citizens Juries c.i.c., a social enterprise that designs and runs citizens' juries in partnership with the Jefferson Center the developers of the citizens' jury method, and funded by the National Institute for Health Research Applied Research Collaboration Greater Manchester (NIHR ARC GM).
- 14 Nuffield Council on Bioethics. The collection, linking and use of data in biomedical research and health care: ethical issues. 2015.
- 15 World Medical Association. Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects. As amended by the WMA General Assembly, Fortaleza, Brazil, October 2013.
- 16 *Coco v AN Clark (Engineers) Ltd* [1968] F.S.R. 415[5].
- 17 This means that the approach and requirements of the law of confidentiality, the interrelated tort of misuse of private information, and an action against a public authority for infringement of human rights, should be largely equivalent where the information is of a private nature (e.g. medical information).

- 18 *Campbell v Mirror Group Newspapers* [2004] UKHL 22, [2004] 2 AC 457, [14]; *Mosley v News Group Newspapers Ltd* [2008] EWHC 1777 (QB), [2008] EMLR 20, [7].
- 19 *Campbell v Mirror Group Newspapers* [2004] UKHL 22, [2004] 2 AC 457, [95]-[96].
- 20 Department of Health. Confidentiality: NHS Code of Practice. 2003.
- 21 General Medical Council. Confidentiality: Good practice in handling patient information. 2017.
- 22 Chico V, Taylor MJ. Using and disclosing confidential patient information and the English common law: What are the information requirements of a valid consent? *Medical Law Review*. 2017; 26(1): 51-72.
- 23 *ibid*.
- 24 General Medical Council. Confidentiality: Good practice in handling patient information. 2017, 13.
- 25 Taylor MJ, Wilson J. Reasonable Expectations of Privacy and Disclosure of Health Data. *Medical Law Review*. 2019; 27(3): 432-460.
- 26 *ibid*.
- 27 General Medical Council. Confidentiality: Good practice in handling patient information. 2017, 14.
- 28 *Campbell v Mirror Group Newspapers* [2004] UKHL 22, [2004] 2 AC 457.
- 29 *Mosley v News Group Newspapers Ltd* [2008] EWHC 1777 (QB), [2008] EMLR 20, [7]; *Trimingham v Associated Newspapers Ltd* [2012] EWHC 1296 (QB), [2012] 4 All ER 717; Taylor MJ, Wilson J. Reasonable Expectations of Privacy and Disclosure of Health Data. *Medical Law Review*. 2019; 27(3): 459.
- 30 Taylor MJ, Wilson J. Reasonable Expectations of Privacy and Disclosure of Health Data. *Medical Law Review*. 2019; 27(3): 459.
- 31 National Data Guardian. Sharing data in line with patients' reasonable expectations. November 2017.
- 32 General Medical Council. Confidentiality: Good practice in handling patient information. 2017, 106.
- 33 Enabled by the National Health Service Act 2006 (as amended), s 251.
- 34 NHS Act 2006, s251(12).
- 35 *ibid*, s251(10).
- 36 Regulation 4, Control of Patient Data Regulations.
- 37 Care Act 2014, Schedule 7 para 8 (1)(a).
- 38 Medical Research Council, NHS Health Research Authority. Is my study research? [online]. Available from: <http://www.hra-decisiontools.org.uk/research/>
- 39 Health Research Authority. CAG pre-application checklist. Available from: https://s3.eu-west-2.amazonaws.com/www.hra.nhs.uk/media/documents/CAG_pre-application_checklist.pdf
- 40 Regulation 3(3), Control of Patient Data Regulations.
- 41 Department of Health and Social Care. Coronavirus (COVID-19): notification to organisations to share information. 1 April 2020. [Updated 10 September 2021]
- 42 Department of Health and Social Care. Coronavirus (COVID-19): notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 – general. 27 August 2021. [Updated 10 September 2021]

- 43 Department of Health and Social Care. Coronavirus (COVID-19): notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 – NHSE, NHSI. 27 August 2021. [Updated 10 September 2021]
- 44 Department of Health and Social Care. Coronavirus (COVID-19): notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002, which were made under sections 60 (now section 251 of the NHS Act 2006) and 64 of the Health and Social Care Act 2001 – Biobank. 27 August 2021. [Updated 10 September 2021]
- 45 Department of Health and Social Care. Coronavirus (COVID-19): notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 – NHS Digital. 27 August 2021. [Updated 10 September 2021]
- 46 Department of Health and Social Care. Coronavirus (COVID-19): notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 – general. 27 August 2021. [Updated 10 September 2021]
- 47 NHS Act 2006, s251(7) (as amended).
- 48 The Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. 2012, 1-10.
- 49 Rumbold JMM, Pierscionek B. The effect of the general data protection regulation on medical research. *Journal of Medical Internet Research*. 2017; 19(2): 1-6. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European Journal of Human Genetics*. 2018; 26(2): 149-156. Donnelly M, McDonagh M. Health research, consent and the GDPR exemption. *European Journal of Health Law*. 2019; 26(2): 97-119. Quinn P. The Anonymisation of Research Data — A Pyrrhic Victory for Privacy that Should Not Be Pushed Too Hard by the EU Data Protection Framework? *European Journal of Health Law*. 2017; 24(4): 347-367
- 50 European Data Protection Board. Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) (art. 70.1.b). January 2019, para 20.
- 51 Article 29 Data Protection Working Party. Guidelines on consent under Regulation 2016/679. European Commission. 2018, 12.
- 52 Hallinan D. Broad consent under the GDPR: An optimistic perspective on a bright future. *Life Sciences, Society, and Policy*. 2020; 16(1): 1-18.
- 53 Dove ES, & Chen J. Should consent for data processing be privileged in health research? A comparative legal analysis. *International Data Privacy Law*. 2020; 10(2): 117–131.
- 54 Mitchell C, Ordish J, Johnson E, Brigden T, Hall A. The GDPR and Genomic Data: PHG Foundation report on the impact of the General Data Protection Regulation and Data Protection Act 2018 on Regulating Genomic Technologies in Healthcare for the Information Commissioner's Office. PHG Foundation. 2020. p65.
- 55 NHS Health Research Authority. Consent in research. [Last updated 19 April 2018]. Available from: <https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/what-law-says/consent-research/>
- 56 Mitchell C, Ordish J, Johnson E et al. The GDPR and Genomic Data. PHG Foundation. May 2020, 68.

- 57 NHS Health Research Authority. UK Policy Framework for Health and Social Care Research. [Last updated 30 April 2021] Section 3.1.
- 58 *ibid.*
- 59 NHS Health Research Authority Research Ethics Service. Defining Research Table. October 2017. Available from: http://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2017-1.pdf
- 60 National Data Guardian for Health and Care. Review of Data Security, Consent and Opt-Outs. 2016. 3.2.21.
- 61 NHS Digital. National Data Opt-out Operational Policy Guidance Document. Version 4.0. February 2020. 1–53.
- 62 NHS Digital. Read more about the collection and conversion of type 2 opt-outs. [Last updated 7 March 2018]. Available from: <https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/how-we-look-after-your-health-and-care-information/your-information-choices/how-opt-outs-work>
- 63 NHS Digital. Individual care and research and planning uses of data. [Last updated 2 August 2019]. Available from: <https://digital.nhs.uk/services/national-data-opt-out/understanding-the-national-data-opt-out/individual-care-and-research-and-planning-uses-of-data>
- 64 NHS Digital. Data Provision Notices (DPNs). Available from: <https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/directions-and-data-provision-notice/data-provision-notice-dpns>
- 65 NHS Digital. National Data Opt-out Operational Policy Guidance Document. Version 4.0. February 2020. 1–53.
- 66 NHS Digital. National Data Opt-out Operational Policy Guidance Document. Version 4.0. February 2020. Chapter 5. 21-23.
- 67 NHS Digital. [MI] National Data Opt-out, September 2021. 9 September 2021. Available from: <https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/september-2021>
- 68 NHS Digital. Patient data and confidential patient information. [Last updated 2 August 2019]. Available from: <https://digital.nhs.uk/services/national-data-opt-out/understanding-the-national-data-opt-out/confidential-patient-information>
- 69 NHS Digital. Understanding the health and care information we collect. [Last updated 7 May 2021] [accessed 2 June 2021]; Available from: <https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/how-we-look-after-your-health-and-care-information/understanding-the-health-and-care-information-we-collect>
- 70 NHS Digital. Data Access Request Service (DARS). [Last updated 25 October 2021]. Available from: <https://digital.nhs.uk/services/data-access-request-service-dars>
- 71 NHS Digital. Data Access Request Service (DARS) products and services. [Last updated 16 September 2021]. Available from: <https://digital.nhs.uk/services/data-access-request-service-dars/dars-products-and-services>
- 72 NHS Digital. NHS Digital transparency notice: GPES data for pandemic planning and research (COVID-19). 13 May 2020. [Last updated 5 August 2021]. Available from:

<https://digital.nhs.uk/coronavirus/gpes-data-for-pandemic-planning-and-research/transparency-notice-gpes-data-for-pandemic-planning-and-research-covid-19>

- 73 NHS Digital. SGSS and CHES data. [Last updated 27 April 2021]. Available from: <https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/directions-and-data-provision-notice/data-provision-notice-dpns/sgss-and-chess-data#legal-basis-for-provision-of-data>
- 74 Public Health England. PHE letter to NHS Trusts Re: COVID-19 Hospitalisation in England Surveillance System (CHES) – daily reporting. 11 March 2020. Available from: <https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/phe-letter-to-trusts-re-daily-covid-19-hospital-surveillance-11-march-2020.pdf>
- 75 Stokes P. The ‘Five Safes’ – Data Privacy at ONS. Office for National Statistics. 27 January 2017. Available from: <https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>
- 76 NHS. NHS COVID-19 Data Store. 2021 [accessed 2 June 2021]. Available from: <https://www.england.nhs.uk/contact-us/privacy-notice/how-we-use-your-information/covid-19-response/nhs-covid-19-data-store/>.
- 77 NHS. NHS England’s Privacy Notice. Available from: <https://www.england.nhs.uk/contact-us/privacy-notice/how-we-use-your-information/covid-19-response/nhs-covid-19-data-store/>
- 78 NHS. National COVID-19 and Flu Vaccination Programmes. Available from: <https://www.england.nhs.uk/contact-us/privacy-notice/national-flu-vaccination-programme/>
- 79 Health Data Research UK. COVID-19 National Core Studies. Available from: <https://www.hdr.uk.ac.uk/covid-19/covid-19-national-core-studies/>
- 80 Health Data Research UK. About Data and Connectivity. Available from: <https://www.hdr.uk.ac.uk/covid-19-data-and-connectivity/about-data-and-connectivity/>
- 81 Health Data Research UK. Our work to help SAGE. Available from: <https://www.hdr.uk.ac.uk/covid-19/our-work-to-help-sage/>
- 82 Health Data Research UK. Patient and Public Involvement and Engagement. Available from: <https://www.hdr.uk.ac.uk/what-is-health-data-research/patient-and-public-involvement-and-engagement/>
- 83 Health Data Research UK Innovation Gateway. Gateway to health data and tools for research. Available from: <https://www.healthdatagateway.org/>
- 84 NHS Digital. Trusted Research Environment service for England. [Last updated 28 September 2021]. Available from: <https://digital.nhs.uk/coronavirus/coronavirus-data-services-updates/trusted-research-environment-service-for-england>
- 85 Health Data Research UK. How do we collect, store and manage data about people’s SARS-CoV-2 samples? 26 January 2021 [accessed 3 June 2021]. Available from: <https://www.hdr.uk.ac.uk/news/how-do-we-collect-store-and-manage-data-about-peoples-sars-cov-2-samples/>
- 86 Health Europa. COVID-19 Genomics UK Consortium: sequencing COVID-19. 12 May 2021. Available from: <https://www.healtheuropa.eu/sequencing-covid-19/108229/>
- 87 COVID-19 Genomics UK Consortium (COG-UK). An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe*. 2020; 1(3): p.e99-e100.

- 88 COVID-19 Genomics UK Consortium. Public Data and Analysis. Available from: <https://www.cogconsortium.uk/tools-analysis/public-data-analysis-2/>
- 89 COVID-19 Genomics UK Consortium. New SARS-CoV-2 variant & other mutations under review. 19 January 2021. Available from: <https://www.cogconsortium.uk/the-new-sars-cov-2-variant-and-other-mutations-under-review-by-cog-uk/>
- 90 UK Research and Innovation. COVID-19 Genomics UK Consortium (COG-UK): Publications. Available from: <https://gtr.ukri.org/project/1BFEFB89-4275-467C-80AC-B4D62347920A#/tabOverview>
- 91 COVID-19 Genomics UK Consortium. How SARS-CoV-2 sequencing is becoming a national service. 23 July 2021. Available from: <https://www.cogconsortium.uk/how-sars-cov-2-sequencing-is-becoming-a-national-service/>
- 92 Peacock S. What next for COG-UK? COVID-19 Genomics UK Consortium. 20 April 2021. Available from: <https://www.cogconsortium.uk/what-next-for-cog-uk/>
- 93 Peacock S. Why COG-UK is joining the UK Health Data Research Alliance. COVID-19 Genomics UK Consortium. 8 July 2021. Available from: <https://www.cogconsortium.uk/why-cog-uk-is-joining-the-uk-health-data-research-alliance/>
- 94 Public Health England has been abolished and its functions transferred to other agencies taking on different aspects of public health with effect from 1st October 2021: the UK Health Security Agency is a new public health body with an emphasis on health protection and security; [<https://www.gov.uk/government/news/uk-health-security-agency-launches-with-a-relentless-focus-on-keeping-the-nation-safe>]; the Office for Health Improvement and Disparities is responsible for the prevention agenda.
- 95 Public Health England. Sources of COVID-19 surveillance systems. [Last updated 29 July 2021]. Available from: <https://www.gov.uk/government/publications/national-covid-19-surveillance-reports/sources-of-covid-19-systems>
- 96 Public Health England. Laboratory reporting to Public Health England: A guide for diagnostic laboratories. October 2020.
- 97 Armstrong J et al. Dynamic linkage of COVID-19 test results between Public Health England's Second Generation Surveillance System and UK Biobank. *Microb Genom*, 2020; 6(7).
- 98 Public Health England. PHE letter to NHS Trusts Re: COVID-19 Hospitalisation in England Surveillance System (CHESS) – daily reporting. 11 March 2020. Available from: <https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/phe-letter-to-trusts-re-daily-covid-19-hospital-surveillance-11-march-2020.pdf>
- 99 Public Health England. Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01 – technical briefing 2. December 2020. [accessed 1 June 2021].
- 100 Public Health England, Imperial College, The University of Edinburgh, The University of Birmingham, The Wellcome Sanger Institute. New evidence on VUI-202012/01 and review of the public health risk assessment. Available from: <https://khub.net/documents/135939561/338928724/New+SARS-COV-2+variant+-+information+and+risk+assessment.pdf/b56d4591-0466-1a18-28dc-010e0fdeef53?t=1608569319930>
- 101 Lopez Bernal J et al. Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. *medRxiv*. 2021.
- 102 GenOMICC. Available from: <https://genomicc.org/>

- 103 COVID-19 Genomics UK (COG-UK) consortium. Explainer. Available from:
<https://www.cogconsortium.uk/the-cog-uk-project-hospital-onset-covid-19-infections-hoci-study/>
- 104 University of Portsmouth. Sequencing and Tracking of Phylogeny in COVID-19. Available from:
<https://researchportal.port.ac.uk/en/projects/sequencing-and-tracking-of-phylogeny-in-covid-19>
- 105 UK Health Security Agency. The SIREN study: answering the big questions. Available from:
<https://ukhsa.blog.gov.uk/2021/03/11/the-siren-study-answering-the-big-questions/>
- 106 UK Health Security Agency. SIREN: One year on. Available from:
<https://ukhsa.blog.gov.uk/2021/06/18/siren-one-year-on/>
- 107 UK Biobank. About us. Available from: <https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us>
- 108 UK Biobank. COVID-19 hub. Available from: <https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/covid-19-hub>
- 109 Torjesen I. Covid-19 will become endemic but with decreased potency over time, scientists believe. *BMJ*. 2021; 372: p. n494
- 110 Martin MA, VanInsberghe D, and Koelle K. Insights from SARS-CoV-2 sequences. *Science*. 2021; 371(6528): p.466-467.
- 111 Andrews SM and Rowland-Jones S. Recent advances in understanding HIV evolution. *F1000Research*. 2017; 6: p. 597.
- 112 Volz E et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell*. 2021; 184(1): p.64-75 e11.
- 113 Volz E et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. 2021; 593(7858): p.266-269.
- 114 Xie X et al. Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. *bioRxiv*. 2021.
- 115 WHO Global Influenza programme & World Health Organisation. Pandemic Influenza Preparedness and Response: a WHO guidance Document. Geneva. 2009.
- 116 Government Digital Service. Data Ethics Framework. June 2018 [updated September 2020].
- 117 Understanding Patient Data, Ada Lovelace Institute. Foundations of Fairness: where next for NHS health data partnerships? March 2020.
- 118 Understanding Patient Data. Trustworthy use of GP data: what must happen now? 15 June 2021. Available from: <https://understandingpatientdata.org.uk/news/trustworthy-use-gp-data-what-must-happen-now>
- 119 Oswald M, Lavery L. Data Sharing in a Pandemic: Three Citizens' Juries – Juries Report. National Institute for Health Research, NHSX, National Data Guardian. May 2021.
- 120 Understanding Patient Data. Health data and Covid-19: taking action while maintaining trust. 17 April 2020. Available from: <https://understandingpatientdata.org.uk/news/health-data-and-covid-19-taking-action-while-maintaining-trust>
- 121 Understanding Patient Data, Ada Lovelace Institute. Foundations of Fairness: where next for NHS health data partnerships? March 2020.

- 122 O'Dowd A. New GP patient data extraction scheme raises concern. *BMJ (Clinical research ed)*. 2021; 373: n1389
- 123 NHS Digital. GP Data for Planning and Research: Letter from Parliamentary Under Secretary of State for Health and Social Care to general practices in England – 19 July 2021. [Last updated 22 July 2021]. Available from: <https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research/secretary-of-state-letter-to-general-practice>
- 124 O'Neill O, What we don't understand about trust. TEDxHousesOfParliament
- 125 Understanding Patient Data. What we mean by trustworthy use of patient data. Available from: <https://understandingpatientdata.org.uk/what-we-mean-trustworthy-use-patient-data>
- 126 Ada Lovelace Institute. Participatory data stewardship. September 2021.
- 127 Guillemin M et al. Do Research Participants Trust Researchers or Their Institution? *Journal of Empirical Research on Human Research Ethics*. 2018; 13(3): 285-294.
- 128 Ipsos MORI Social Research Institute. The one-way mirror: public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. 2016.
- 129 Ipsos MORI. Public attitudes to the use and sharing of their data. Research for the Royal Statistical Society by Ipsos MORI. July 2014.
- 130 National Data Guardian for Health and Care. Review of Data Security, Consent and Opt-Outs. June 2016.
- 131 Wellcome Trust. Public attitudes to commercial access to health data. March 2016. Available from: <https://wellcome.org/sites/default/files/public-attitudes-to-commercial-access-to-health-data-summary-wellcome-mar16.pdf>
- 132 Ghafur S et al. Public perceptions on data sharing: Key insights from the UK and the USA. *The Lancet Digital Health*. 2020; 2(9).
- 133 Ipsos MORI Social Research Institute. Dialogue on Data. Report prepared for the Office for National Statistics and the Economic & Social Research Council. 2014.
- 134 Ipsos MORI Social Research Institute. The one-way mirror: public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. 2016.
- 135 Ibid.
- 136 Ibid.
- 137 Wellcome Trust. Public attitudes to commercial access to health data. March 2016. Available from: <https://wellcome.org/sites/default/files/public-attitudes-to-commercial-access-to-health-data-summary-wellcome-mar16.pdf>
- 138 Oswald M, Lavery L. Data Sharing in a Pandemic: Three Citizens' Juries - Juries Report. National Institute for Health Research, NHSX, National Data Guardian. May 2021.
- 139 National Data Guardian. Press release: Polling indicates growing public understanding about importance of using health and care data. October 2020. Available from: <https://www.gov.uk/government/news/polling-indicates-growing-public-understanding-about-importance-of-using-health-and-care-data>

- 140 NHS Digital. National Data Opt-out, June 2021. Available from <https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/june-2021>
- 141 NHS Digital. National Data Opt-out, July 2021. Available from <https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/july-2021>
- 142 Bharti N et al. Public Trust, Deliberative Engagement and Health Data Projects: Beyond Legal Provisions. *Engaging Science, Technology, and Society*. 2021; 7(1): 125–133.
- 143 House of Commons Public Administration and Constitutional Affairs Committee. Government transparency and accountability during Covid 19: The data underpinning decisions. Eighth report of session 2019-21. March 2021. Available from: <https://committees.parliament.uk/publications/5076/documents/50285/default/>
- 144 Understanding Patient Data, Ada Lovelace Institute. Foundations of Fairness: where next for NHS health data partnerships? March 2020.
- 145 Bharti N et al. Public Trust, Deliberative Engagement and Health Data Projects: Beyond Legal Provisions. *Engaging Science, Technology, and Society*. 2021; 7(1): 125–133.
- 146 McCormack P et al. 'You should at least ask'. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. *Eur J Hum Genet*. 2016; 24(10): 1403–8.
- 147 Understanding Patient Data, Involve and Carnegie UK Trust. Data for Public Benefit: Balancing the risks and benefits of data sharing. April 2018.
- 148 Ipsos MORI Social Research Institute. The one-way mirror: public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. 2016.
- 149 Ipsos MORI. A public dialogue on genomic medicine: time for a new social contract? Genomics England, UK Research and Innovation & Sciencewise. April 2019.
- 150 National Data Guardian. Press release: Polling indicates growing public understanding about importance of using health and care data. October 2020. Available from: <https://www.gov.uk/government/news/polling-indicates-growing-public-understanding-about-importance-of-using-health-and-care-data>
- 151 Understanding Patient Data, Involve and Carnegie UK Trust. Data for Public Benefit: Balancing the risks and benefits of data sharing. April 2018.
- 152 Howe N, Giles E, Newbury-Birch D. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *J Health Serv Res Policy*. 2018; 23(2): 123–33
- 153 Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. *Expert Rev Mol Diagn*. 2014; 14(8): 1053–65
- 154 Ipsos MORI Social Research Institute. The one-way mirror: public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. 2016.
- 155 Ada Lovelace Institute. The data divide: public attitudes to tackling social and health inequalities in the COVID-19 pandemic and beyond. March 2021.
- 156 Understanding Patient Data. Why an opt-out rather than an opt-in or consent. May 2018. Available from: <https://understandingpatientdata.org.uk/news/why-an-opt-out>
- 157 Ibid.

- 158 Ibid.
- 159 National Data Guardian for Health and Care. Review of Data Security, Consent and Opt-Outs. June 2016. Citing an Interview with representative from national patient representative charity National Voices 1 December 2015.
- 160 Ipsos MORI Social Research Institute. The one-way mirror: public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. 2016.
- 161 The King's Fund. Using data in the NHS: the implications of the opt-out and GDPR [online]. May 2018. Available from: <https://www.kingsfund.org.uk/publications/using-data-nhs-gdpr>
- 162 Understanding Patient Data. Putting the Trust in Trusted Research Environments. 28 July 2020. Available from: <https://understandingpatientdata.org.uk/news/putting-trust-trusted-research-environments>
- 163 NHS Health Research Authority. UK Policy Framework for Health and Social Care Research. [Last updated 30 April 2021].
- 164 Understanding Patient Data, Ada Lovelace Institute. Foundations of Fairness: where next for NHS health data partnerships? March 2020.
- 165 Ada Lovelace Institute. Participatory data stewardship. September 2021.
- 166 Understanding Patient Data, Ada Lovelace Institute. Foundations of Fairness: where next for NHS health data partnerships? March 2020.
- 167 Department of Health and Social Care. Annual Report of the Chief Medical Officer 2016: Generation Genome. 2017.
- 168 Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why *care.data* ran into trouble. *Journal of Medical Ethics*. 2015; 41: 404-409.
- 169 Oswald M, Lavery L. Data Sharing in a Pandemic: Three Citizens' Juries - Juries Report. National Institute for Health Research, NHSX, National Data Guardian. May 2021
- 170 Ibid.
- 171 Ada Lovelace Institute. Participatory data stewardship. September 2021.
- 172 Torjesen I. Covid-19 will become endemic but with decreased potency over time, scientists believe. *BMJ*. 2021; 372: p.n494.
- 173 NHS Health Research Authority. COPI notice transition applications. [Last updated 6 October 2021.] Available from: <https://www.hra.nhs.uk/covid-19-research/guidance-using-patient-data/copi-notice-transition-applications/>
- 174 Lintern S. Plans to hand over NHS data to police sparks warning from government advisor. The Independent. [online]. 11 October 2021.
- 175 OpenSAFELY. Available from: <https://www.opensafely.org/>
- 176 NHS Digital. GP Data for Planning and Research: Letter from Parliamentary Under Secretary of State for Health and Social Care to general practices in England – 19 July 2021. [Last updated 22 July 2021]. Available from: <https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research/secretary-of-state-letter-to-general-practice>

- 177 Understanding Patient Data. Listening to patients and publics: a mix of methods, one coherent approach. 11 March 2019. Available from: <https://understandingpatientdata.org.uk/news/listening-to-patients>
- 178 Department for Digital, Culture, Media & Sport. Data: a new direction. 10 September 2021. Available from: <https://www.gov.uk/government/consultations/data-a-new-direction>
- 179 Department for Digital, Culture, Media & Sport. National Data Strategy. [Last updated 9 December 2020].
- 180 Department of Health and Social Care. Data saves lives: reshaping health and social care with data (draft). 22 June 2021. [Last updated 8 September 2021].
- 181 Department of Health and Social Care. Annual Report of the Chief Medical Officer 2016: Generation Genome. 2017. Chapter 16.
- 182 Legal Mechanisms for Data Stewardship working group. Exploring legal mechanisms for data stewardship. UK AI Council & Ada Lovelace Institute. 4th March 2021.
- 183 COVID-19 Genomics UK Consortium. Expanding the COG-UK research portfolio to increase impact from genome data. 24 June 2021. Available from: <https://www.cogconsortium.uk/expanding-the-cog-uk-research-portfolio-to-increase-impact-from-genome-data/>
- 184 Baillie JK et al. GenOMICC study protocol. Genetics of Mortality In Critical Care Consortium (GenOMICC). 24 April 2020. Available from: <https://genomicc.org/protocol/>
- 185 Pairó-Castineira E et al. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021; 591(7848): p. 92-98.
- 186 Randomised Evaluation of COVID-19 Therapy (RECOVERY). Baricitinib to be investigated as a possible treatment for COVID-19 in the RECOVERY trial. 2 February 2021. Available from: <https://www.recoverytrial.net/news/baricitinib-to-be-investigated-as-a-possible-treatment-for-covid-19-in-the-recovery-trial>
- 187 Blackstone J et al. Protocol for the COG-UK hospital onset COVID-19 infection (HOCI) multicentre interventional clinical study: evaluating the efficacy of rapid genome sequencing of SARS-CoV-2 in limiting the spread of COVID-19 in United Kingdom NHS hospitals. *medRxiv*. 2021.
- 188 Stirrup O et al. Rapid feedback on hospital onset SARS-CoV-2 infections combining epidemiological and sequencing data. *eLife*. 2021; 10: e65828.
- 189 SARS-CoV-2 Immunity & Reinfection Evaluation (SIREN). Privacy Information. V2/3.0. 17 June 2021. Available from: <https://snapsurvey.phe.org.uk/siren/data-privacy-notice.pdf>
- 190 Hall VJ. et al. SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). *Lancet*. 2021; **397**(10283): p.1459-1469.
- 191 UK Biobank. Primary Care Linked Data. Version 1.0. September 2019. Available from: https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/primary_care_data.pdf
- 192 UK Biobank. Approved Research. Available from: <https://www.ukbiobank.ac.uk/enable-your-research/approved-research?page=1&terms=genetic+covid#articles>. A search for “Genetic COVID” brought up 13 studies active at the time of access [Accessed 31 August 2021]

- 193 Zhu Z et al. Association of obesity and its genetic predisposition with the risk of severe COVID-19: Analysis of population-based cohort data. *Metabolism*. 2020; 112: p.154345.
- 194 Zhu Z et al. Association of asthma and its genetic predisposition with the risk of severe COVID-19. *J Allergy Clin Immunol*. 2020; 146(2): p.327-329 e4.
- 195 Kuo CL et al. ApoE e4e4 Genotype and Mortality With COVID-19 in UK Biobank. *J Gerontol A Biol Sci Med Sci*. 2020; 75(9): p.1801-1803.
- 196 Dite GS, Murphy NM, and Allman R. An integrated clinical and genetic model for predicting risk of severe COVID-19: A population-based case-control study. *PLoS One*. 2021; 16(2); p.e0247205.
- 197 Toh C and Brody JP. Evaluation of a genetic risk score for severity of COVID-19 using human chromosomal-scale length variation. *Hum Genomics*. 2020; 14(1): p.36.
- 198 Zhu Z et al. Association of obesity and its genetic predisposition with the risk of severe COVID-19: Analysis of population-based cohort data. *Metabolism*. 2020; 112: p.154345.
- 199 Zhu Z et al. Association of asthma and its genetic predisposition with the risk of severe COVID-19. *J Allergy Clin Immunol*. 2020; 146(2): p.327-329 e4.

Appendix 1 – Interviewees

Name of interviewee	Affiliation
Carolina Arevalo	Head of Research Support and Governance, Public Health England
Natalie Banner	Former Lead, Understanding Patient Data
Phil Booth	Coordinator, medConfidential
Judy Breuer	Professor of Virology, Director of the Viral Genomes Unit, University College Hospital
Tony Calland	Chair, Confidentiality Advisory Group, Health Research Authority.
Elizabeth Coates	Head of Research Governance, Public Health England
Paul Comerford	Principle Technology Advisor (Technology and Innovation Service), Information Commissioner's Office
David Evans	Information Governance Policy Lead, NHSX
Mavis Machirori	Senior Researcher, Ada Lovelace Institute
Malcolm Oswald	Honorary Researcher in Law, Centre for Social Ethics and Policy, University of Manchester
Richard Scott	Chief Medical Officer, Genomics England
David Seymour	Executive Director, UK Health Data Research Alliance, Health Data Research UK
Jonathan Sellors	Legal Counsel and Company Secretary, UK Biobank and UK Biocentre
Mark Taylor	Associate Professor in Health Law and Regulation, University of Melbourne

Appendix 2 – Exemplar research studies

This appendix provides more information about the exemplar research studies cited in Section 3 of the report. All these involve linkage of genomic and health data, and cite reliance on COPI notices.

1. GenOMICC

Details of study: The Genetics of Mortality In Critical Care (GenOMICC) consortium is a global community of doctors and scientists trying to understand and treat critical illness led by the University of Edinburgh. It was set up in 2016 to sequence the genomes of patients with serious diseases, primarily flu and other emerging diseases with the aim to help understand how genetics may influence disease outcomes and severity.

In the UK, the GenOMICC consortium is working with Genomics England and COG-UK to understand the role of genetics in the risk of developing severe COVID 19. Genomics England are performing the sequencing of patient genomes as well as linking this data with other health datasets from a variety of sources. Participants will be recruited on the basis that they were previously healthy patients who are currently or were previously in intensive care with severe COVID-19 disease (20,000) as well as 15,000 individuals who experienced mild or moderate symptoms. Analyses will be done on an iterative basis, so initial results will be on a smaller cohort of patients as their genomes are sequenced, and then repeated as the numbers of people who join the programme grow. These cases will be compared with matched controls selected from the 100,000 Genomes Project database.

Data linkage: Patient genome data is linked with data from NHS Digital (mortality data, hospital episode statistics, Emergency Care Data Sets, mental health, cancer registration, diagnostic imaging dataset (no images), patient reported outcome measures, secondary uses dataset), public health agency data (COVID-19 test results), ICNARC (health data from intensive care) and ISARIC (admission, PMH, ventilation, smoking, outcome data). Individual genomes will also be linked to the virus genome data provided by COG UK Consortium. The University of Cambridge will link SARS-CoV-2 sequence data with human genomes generated by GenOMICC.¹⁸³

COPI notice reliance: Some of the data received by Genomics England from NHS Digital are shared under the COPI notices. Although participants in this project have, by default, provided consent for their data to be used for future research projects, the COPI notices require NHS Digital to share confidential patient information with research organisations, entitled to process this under COPI for COVID-19 purposes.¹⁸⁴

Data access to research community: In addition to the analyses done by the GenOMICC consortium, the data collected will be available in Genomic England's National Genomic Research Library for research and made available to researchers. Summary level data from the study can be accessed broadly for research

purposes. More detailed data will be available to core groups, and to researchers who might contribute to the programme through their Trusted Research Environment.

Potential insights: By comparing the genomes of patients who become seriously ill with COVID 19 to those who experience mild or moderate disease, insights can be derived about the influence of genetics independent of other known risk factors. In addition, linking patient genomes with the genomes of the virus the patient is infected with can reveal any virus-host interactions. Results from this study could help to identify individuals at higher or lower risk based on their genetics.

In the longer-term, the data could also help to inform personalisation of treatments or vaccines. For example, there may be an influence of genotype on sustained response to specific vaccines. Whilst the potential insights from this study are directly relevant to SARS-CoV-2 they could be more broadly applicable to other infectious diseases.

Published findings: GenOMICC published interim findings from a genome-wide association study (GWAS) in November 2020 from 2244 critically ill COVID-19 patients. The regions identified to be associated with severe disease through the GWAS were mapped to the genes they are in or near to. These genes were investigated for their role in disease pathogenesis.

A number of significant gene-disease associations were identified and replicated which have had implications for prognosis, treatment and management of patients with COVID-19. For example, that low expression of IFNAR2, or high expression of TYK2, are associated with life-threatening disease; and transcriptome-wide association in lung tissue revealed that high expression of the monocyte–macrophage chemotactic receptor CCR2 is associated with severe COVID-19.¹⁸⁵ This association has led to the discovery that the TYK2 variant, resulting in higher expression of the protein involved in inflammation, could be targeted by baricitinib, a drug currently used for rheumatoid arthritis. The randomised controlled trial, Adaptive Covid-19 Treatment Trial, has shown promising results for this drug and it has subsequently been incorporated into the RECOVERY trial¹⁸⁶ in the UK.

2. Project Hospital-Onset COVID-19 Infections Study

Details of study: Hospital-Onset COVID-19 Infections (HOICI) Study is a project initiated by COG-UK. It is a phase III prospective, interventional, cohort, superiority study to evaluate the benefit of rapid COVID-19 genomic sequencing (the COVID-19 GENOMICS UK project) on infection control in preventing the spread of the virus in United Kingdom NHS hospitals. The study began in October 2020 and it will identify not only whether rapid viral sequencing is useful for patient management, but how time-critical this might be.

2,000 patients who test positive for COVID-19 and who have hospital-onset infection will have their sample sequenced as part of the wider COG-UK sequencing project. Findings will be fed back to Infection Prevention Control (IPC) teams within hospitals to inform their actions. Fourteen NHS hospital Trusts/Health Boards across England and Scotland are participating in study.

The sequencing report will bridge a knowledge gap between bioinformatics analysis and infection prevention and control teams to show where patients fit with other infections in a hospital outbreak.

Data linkage: Viral genomic data is linked with patient data and ward level data (i.e. which wards patients and staff have been on).

COPI notice reliance: Consent for participant (both patient and healthcare workers) involvement will not be sought for COG-UK HOCl study. Instead, this project relies on COPI notices. This approach was reviewed and approved by a Research Ethics Committee.¹⁸⁷

Data access to research community: The terms of the funding requires the COG-UK HOCl study dataset to be shared on CSDR (clinicalstudydatarequest.com) or an equivalent data sharing platform, within 6 months of public reporting of results so that the data may be reused by other researchers.

Potential insights: This study aims to evaluate the benefit of COVID-19 genomic sequencing and return of data report on infection control in preventing the spread of the virus in UK NHS hospitals. It will also determine if rapidly sequencing and producing a report (within 48 hours after receipt of sample) delivers more impact than standard time to sequencing report (5 – 10 days). Potential benefits of genomic sequencing of samples from patients with hospital-onset infection include: determining whether the infection is hospital or community acquired, defining the occurrence and transmission location of these infections, and identify previously undetected nosocomial transmission, all of which could help to implement strategies to reduce the incidence rate of IPC-defined HOCl.

Other insights that will be gained from this study include: what factors affect the benefit of sequencing pathogens for IPC, the cost implications, the factors influencing nosocomial spread, association between viral lineages and the biology of disease (e.g. severity and/or viral load).

The results of this study could have wider impacts around future decisions to utilise genome sequencing for other pathogens (e.g. influenza, antimicrobial resistant pathogens, norovirus, *Clostridium difficile* and respiratory syncytial virus) in secondary care settings.

Published findings: Early proof of concept study demonstrated the use of a statistical method and sequence reporting tool combining epidemiological and sequence data to assess the probability that hospital onset COVID-19 infections were acquired in hospital.¹⁸⁸ This initial study formed the basis of the HOCl study. Since results were not reported in real time, and transmission chains were not inferred from the study findings, there were limited impacts on outbreak management, however other studies have attempted to provide more immediate feedback to support public health activities.

3. Sequencing and Tracking of Phylogeny in COVID-19

Details of study: The University of Portsmouth are working with COG-UK Consortium and other groups across the globe, to identify distinct clusters of the virus as the pandemic progresses via the Sequencing

and Tracking Of Phylogeny in COVID-19 (STOP COVID-19) study. The main aim is to generate a database of viral RNA sequences for SARS-CoV-2 within the Wessex region.

They are using genomic sequencing to assemble viral genomes from patients who have presented to Portsmouth Hospital NHS Trust (PHT) in the Wessex region with symptoms of COVID-19. Whole genome sequences will be compared using phylogenetic analysis to identify the spread of the virus within the local area, will be analysed in the context of anonymised patient level data to look for trends in the adaptation of the virus, and will be compared with a global database of such sequences to help develop global maps of transmission.

Data linkage: whole genome sequencing data of the virus is linked with what is described as anonymous patient information.

COPI notice reliance: The study is listed on the Health Research Authority website with a note that the study relies on COPI notice.

Data access to research community: N/A

Potential insights: By analysing the nature and speed of the changes in viral genomic data, these data can be used to identify potential transmission clusters, to identify the spread of the virus within the local region, and can be compared with other sequences generated globally to help develop global maps of transmission. In addition, the effects of different strains of the virus on patient outcomes, and how the specific strain of the virus may impact the health care of the participant is being investigated in this study.

Understanding the evolution of the virus and its spread across the globe will help researchers to (1) predict the future spread of the virus (2) estimate the number of worldwide cases, and (3) aid in the development of epidemiological models for estimating a potential end point to the pandemic crisis. In addition, the ability to track mutations in real time allows researchers access to a large body of data to explore in order to identify potential targets for cures and vaccines.

Published findings: Data from this study feeds into COG-UK wider surveillance activities and is incorporated into Microreact, an online dashboard displaying SARS-CoV-2 lineages.

4. SIREN study

Details of study: The SIREN study is a cohort of 44,546 NHS workers over 135 sites in the UK (25,693 in England) who receive testing for COVID-19 by PCR tests every 2 weeks. The study is being conducted by PHE and aims to determine the incidence, characteristics and potential of new infections. Positive samples are sent for genomic sequencing to determine how closely related viruses from different individuals are to each other.

Data linkage: Questionnaire responses and results from COVID-19 PCR tests and antibody tests will be linked to personal identifiable details provided by participants. These datasets may also be linked to

participants' health and care records, for example, the National Immunisation Management System to determine vaccination status or hospitalisation episodes to investigate severity of infection.

COPI notice reliance: This study relies on COPI notices to enable linking test result data to health records.

Data access to research community: Non-identifiable information may be collected, analysed, reported and shared with others within Europe to contribute to research. More information can be reviewed in the SIREN Privacy Statement.¹⁸⁹

Potential insights: The study aims to understand whether prior infection with SARS-CoV2 protects individuals against future infection with the same virus. Regular testing (for the virus and antibodies to the virus) of frontline staff will also help to understand the number of workers who have been infected by SARS-CoV-2 and if there are associations or different responses to infection depending on different factors including age, profession and ethnicity. Regular blood samples to test for antibody levels will reveal how antibody levels change over time and the different types of antibodies that are present. It may also reveal if new variants of SARS-CoV-2 have emerged that are able to evade natural or vaccine induced immunity. For example, the incidence of new infections and reinfections would be expected to rise if a new variant was able to evade immunity.

Published findings: The interim analysis of the primary study objective, to determine whether prior infection is protective against future infections, was published in April 2021.¹⁹⁰ Results showed that a previous history of SARS-CoV-2 infection was associated with an 84% lower risk of infection, with median protective effect observed 7 months following primary infection. These results show that immunity gained from previous infection is protective against further infection in most individuals.

5. UK Biobank

Details: UK Biobank recruited 500,000 people aged between 40-69 years in 2006-2010 from across the UK to provide biological samples (including blood, urine and saliva), detailed information about themselves and agree to share their health data throughout their life. The aim of biobank is to improve the prevention, diagnosis and treatment of a wide range of diseases and conditions.

COPI notice reliance: The COPI notices have resulted in the UK Biobank receiving multiple medical records of some 400,000 participants. These include COVID-19 diagnostic tests, deaths, GP records and hospital episodes – including, for the first time, critical care events. This has facilitated accelerated access to patient records, particularly GP records.¹⁹¹

Data access to research community: The data contained within UK Biobank is a tool for health research and is available to all researchers, whether in universities, charities, government agencies or commercial companies, based in the UK or abroad through the same application process and approval criteria. UK

Biobank is making regularly updated health outcome data available for COVID-19 related research to all researchers worldwide who have approval to use the resource.

Potential insights: Integrating data from the wide range of medical records with the genetic and lifestyle data already available in the resource will enable researchers to identify individuals diagnosed with COVID-19 across the full spectrum of disease severity (i.e. not limited to hospitalised cases). The linkage of COVID-19 test data with participant medical records, genetic and lifestyle data will help researchers to understand the complex interplay between genetics, lifestyle and underlying health conditions on the outcomes, severity and recovery from COVID-19. Linking these datasets also presents opportunities for researching the longer-term health effects of COVID-19 as participants continue to have their health data collected over time.

There are 13 active research studies approved to access Biobank data investigating the role of genetics in COVID-19 severity, susceptibility and recovery to date.¹⁹²

Published findings: There are many published findings using Biobank data relating to COVID-19. Several of these include genetic/ genomic information about participants alongside clinical information including COVID-19 infection status and reported outcomes. Many of these studies attempt to associate risk of severe COVID-19 with genetic predisposition of diseases such as obesity¹⁹³ and asthma,¹⁹⁴ or specific genotypes (e.g. ApoE e4e4).¹⁹⁵ In addition, risk prediction models incorporating genetic risk factors alongside standard clinical details have been developed,^{196 197} providing early indications that genetic susceptibility may be useful for defining risk of individuals.^{198 199}

Appendix 3 – Traverse focus group report

Patient data use in research: Focus Group report

Report by Traverse
July 2021



Introduction

Project background

Context

During the Covid-19 pandemic, the Department of Health and Social Care issued a series of Control of Patient Information (COPI) notices, which allowed confidential patient data to be shared with research organisations for the purpose of Covid-19 research. These notices allowed organisations to bypass the need to ask for patients' consent to use their data, or get the permission of a specialist committee (the Confidentiality Advisory Group). Other safeguards remained in place.

The COPI notices were initially in place until September 2020 and have been extended until September 2021. As we begin to move beyond the Covid-19 pandemic, and the end-date of the COPI notices, the PHG Foundation has been commissioned to conduct a legal and ethical analysis of these changes in patient data sharing to inform future policy. This piece of work, conducted by an independent research organisation, Traverse, forms a part of this review. It explores the views of the public on the changes brought about by the COPI notices, and their views on what should happen next.

Our research focused on two key questions, which formed the basis of the conversation in the focus group:

- What are your views on whether it is okay for confidential data to be collected and used for healthcare and research more freely than usual during the pandemic?
- What do you think should happen when the pandemic is over?

Method

To explore the public's views on the use of patient data in research, Traverse delivered an online focus group in June 2021 with ten participants. The session lasted two hours.

Sample

Participants were recruited from across England through a recruitment agency. The group was recruited to be broadly representative of the UK population, with a range of participants across age, gender, ethnicity, and socio-economic status. Participants were recruited in certain areas, meaning there was a concentration of people living near Greater Manchester and London. All participants were selected on the basis that they had had contact with their healthcare provider in the last 12 months.

We excluded certain roles from our recruitment, for example, people who worked in medical research, as this may have biased the conversation. Participants were provided with a payment to thank them for their time and contribution.



Participant	Gender	Age	Ethnicity
AH	M	48	White British
AH	M	36	Pakistani
BW	M	59	White British
CR	F	36	White British
GP	F	64	White other
JS	M	69	White British
KM	F	31	White British
MH	F	78	Black Caribbean
SS	M	24	Black African
ST	F	70	White British

Approach

A topic guide was developed in collaboration with the PHG Foundation (see annex 3).

In order to gather individual views and compare attitudes to patient data sharing before and after the session, we asked participants to complete two short online questionnaires (see annex 2)

- The first explored their attitudes towards the use of patient data in research, and any potential benefits or risks they could think of. We provided a brief explanation of patient data to help participants understand the subjects being discussed.
- The second repeated the same questions and asked them to choose their favourite option for patient data use post-pandemic.

In the group we asked participants for their unprompted views on the topic. Participants were told about the change in patient data sharing during the Covid-19 pandemic and asked to reflect. They were also asked to consider their preference for patient data sharing regulation moving forwards.

Limitations

The research involved a small self-selecting sample, the people taking part were those who were willing and able to. The views expressed should not be considered representative or generalisable.

On the day of the group, the news of the GP Data for Planning and Research (GPDPR) initiative had recently broken, along with the announcement that the scheme had been deferred. The GPDPR plan caused some confusion for participants, and it was front of mind for some. However, this potentially allowed participants to talk about patient data more freely and in a more concrete manner – due to the similar nature of both topics.



Reading this report

This report presents thematic analysis of findings from the group, divided broadly into the following categories:

- Benefits to data sharing
- Concerns about data sharing
- Increasing trust

The slides from the session, the online questionnaire and a summary of the participant sample can be found in the appendix.

Thematic analysis of findings

Benefits to data sharing

Data from the initial questionnaire showed overall support for the sharing of patient data to help tackle Covid-19, with five participants agreeing and two strongly agreeing with the statement “I am happy for my own patient data to be shared with researchers and the NHS if it helps tackle Covid-19”. A similar picture was reflected when participants were asked about the sharing of patient data to tackle any illness or health condition, and when asked about its use to help public agencies plan and deliver health and care. Based on the second questionnaire, conducted after the focus group, a shift can be observed across responses to all questions about the benefits of data sharing which suggests greater willingness for patient data to be shared in these circumstances. Following the focus group, all participants identified at least one potential benefit of patient data sharing.¹

Participants identified a number of key benefits of the use of patient data in medical research.

First, they saw patient data sharing as **essential for medical treatment** and follow-up, particularly for ease of care, and especially when moving between different services and professionals. One participant mentioned the benefit of data sharing when it came to follow-up appointments after he had cancer.

Participants also thought sharing patient data would help **identify people at risk of disease** so they could be treated earlier.

Participants expressed a lot of positive sentiment towards the use of patient data to benefit the **health of future generations** and work towards **medical advancement**. They quickly recognised the potential for developing new vaccines and treatments, monitoring current treatments: side-effects, effectiveness of treatments and increasing our understanding of science. There was also recognition of scale and scope, with one participant identifying that health trends could be picked up quicker with larger data sets. One participant argued that using genetic data for medical research was essential to do at the moment so it wouldn't be lost in time.

¹ See Appendix for more information



"Every bit of information is valuable for future research." Male, 69, White British

They also saw the clear benefit for anticipating needs in healthcare **and planning for future care**. Reflecting on the events of the past year, participants saw the use of patient data as helpful to predict and avoid future pandemics or other public health crises. In particular, patient data could be shared to support **better resource planning**, by developing our understanding of needs in certain geographical areas, certain conditions or certain demographics. This was seen to allow for better delivery of care as well as anticipating health crises.

"I think the information, the data, that the NHS holds already is vital because, without it, we wouldn't have service planning and health promotion. So if you've got one area of Manchester, say, where there seems to be an awful lot of people having respiratory problems – ignore Covid, it's just in general – then health officials and local authorities need to look at why there are so many people in a specific area." Female, 36, White British

Whilst there was significant convergence of views on the promises and benefits of patient data sharing, there was some disagreement, particularly about whether routinely gathering biometric data could be justified. A couple of participants were very favourable to the idea, as they saw clear societal benefits of the use of this data in research, while others were strongly opposed due to fears surrounding data security and misuse.

Concerns about data sharing

How the data is kept

Upon first reflection, participants assumed that patient data is kept securely and used appropriately, in ways they would approve of. Since most had not thought about the topic before, it had not occurred to them that there may be risks when sharing their data.

"Because it's your personal data, you assume it's being stored and used appropriately." Male, 48, White British

However, on consideration, the group quickly raised concerns around data security, and worries around hacking, data breaches or data loss. These concerns reflected fears and confusion around data sharing and technology in general. Many spoke about the risks of security breaches and misuse by third parties, including for criminal activities. This was also reflected in questionnaire responses before and after the focus group, with most citing data breaches or "data getting into the wrong hands" as their main concern on the topic.



"If anything gets hacked or misused, or my data gets put anywhere I wouldn't be comfortable with that." Male, 36, Pakistani

Even when safeguards are in place, participants were vocal about their desire for information security protocols to be followed to avoid breaches, with no allowances for "cutting corners".

Participants were keen that data be deleted after it was used for its specific purpose, so it would not be used for other purposes or by other actors, or fall "into the wrong hands".

"There is no information about how the information will be stored and used after the studies are completed. Information could get into the wrong hands."
Male, 48, White British

Identification leading to discrimination was identified as a key risk. This was seen as a particular risk for stigmatised conditions, including mental health problems. For instance, participants thought identification could lead to a person losing their job if they were found to have a certain condition. Some were concerned that insurance companies could also use identifiable patient data to charge customers more if they had certain conditions or lifestyles.

"If I go for health insurance and I'm a smoker and they've got that on record they can then use that against me even if I've stopped for five years or something and I don't have to declare it. They can still see that on my record [...] so how it can be used against you I think for me would be one of the things." Female, 31, White British

However, there was variance across the group, and some participants highlighted certain risks more than others. Some were more concerned about some of the risks from the use of patient data than others were. A small number of participants were very favourable to the idea of their data being used in research, so long as it was 'safe' and 'appropriate'. On the other hand, some participants were less favourable, and showed a higher level of concern around the personal risks data sharing may entail.

Who has access to the data?

Participants' level of trust in relation to who has access to data varied greatly from organisation to organisation. As discussed previously, participants were concerned around the use of their data by non-health related organisations. Trust in healthcare-specific bodies was generally higher.

Participants were generally comfortable with the NHS using their data for medical research, since they felt the NHS already held this data on them and were highly trustworthy. They were

also more supportive of the NHS using their patient data without their direct consent than they were for any other organisation.

"I wouldn't mind anything within the NHS... research into cancer, meningitis. Anything that has a cause." Female, 31, White British

Trust in the government's use of data for health-purposes (e.g. service planning) was a little more controversial. Although some participants had fairly high trust in the government, others were more concerned. The failure of certain initiatives throughout the pandemic had eroded trust in the government's management of data. Participants cited the failure of NHS Test and Trace, as well as issues with identifying people on the shielding list for support.

"If it's government-led studies, I mean, I don't even trust them after they've messed up with the track and trace. How are they going to deal with everyone's data? It's a bit iffy for me." Male, 59, White British

Participants were also concerned with the potential for data sharing within government and with organisations they did not approve of. This concern was particularly top-of-mind due to the recent story on the GDPR.

Trust in private or profit-driven companies was particularly low. Organisations such as pharmaceutical companies, insurance companies or "American private healthcare firms" were cited as examples of organisations that people would not want to have access to their patient data. Participants worried that these organisations would sell that data on for money or that large corporations would use personal data, as one participant put it, as a "bargaining tool". Participants also didn't trust the motivations of these organisations and doubted that they would work towards the uses of patient data they approved of (medical advancement or planning for the future).

There was also concern around the potential of sharing patient data with non-health bodies, such as the criminal justice system, the Department of Work and Pensions or other departments of government. This was linked to fears of identification and discrimination.

Why is the data needed?

Participants were broadly in favour of any uses that would have a clear public benefit. They struggled to define this in terms of specifics, but these generally aligned with what they saw as the main benefits of using patient data for medical research: supporting better individualised medical care, monitoring population health, anticipating future healthcare needs and making scientific advancements.

Less acceptable were 'commercial' uses which would allow for people other than the 'public' to benefit from using patient data.



Participants were also concerned with function creep or change of purpose – they wanted to be sure that their personal data would only be used for the purpose they had agreed. They were concerned that their patient data would be used for non-health uses without their knowledge or consent.

One participant hypothesised that the government may be planning on using care records as a form of indirect Covid-19 vaccine passport. Although this was a minority view and a context-specific example, it is worth noting as it illustrates participants' concern that their data could be used to restrict their independence and individual liberties.

However, in general participants were less concerned about **why** patient data was used than **who** it was used by – which suggests participants assumed that certain organisations would have their best interests at heart or hold values that aligned with their own.

What data is used?

Another key area of concern to participants was the type and amount of data shared.

Generally, there was a high level of concern around blanket data sharing, and that researchers would use all of the data they could gather about one person out of convenience or 'because they can', even if that information wasn't directly related to the research. Instead, participants wanted the data used to be directly relevant to the study in question.

"I don't trust where blanket, wholesale information is given. If it's for a specific purpose, I have no problem." Female, 36, White British

Participants preferred that the data be specific and directly related to medical research. For example, one of the main reflections on the use-cases was that each of the studies had used very specific data which directly aligned with the purpose of the research.

There was reluctance around demographic information being used, as well as any identifying information. Participants preferred researchers to only have access to group or population-level statistics rather than individual data, as they saw this as safer (in relation to identifiability) and more relevant to research.

Anonymisation was another key requirement participants identified. However, this raised a couple of questions and concerns. There was some disagreement about the potential of de-anonymising data, and the potential risks of identification, particularly around stigma and employment. This topic did cause confusion for some, with participants questioning how consent could be sought if researchers were using anonymised datasets. This then also raised the question of whether consent would be sought if that data was re-identified.

Increasing trust

What impacted trust levels?

Some of the participants' views were influenced by **cultural norms**. Their views on the acceptability of data sharing (both in general and specifically of patient data) depended on what they had been exposed to. One participant originated from a country where routine



biometric data collection was common. This impacted that person's views throughout the session, as they viewed data collection and sharing as 'normal'. On the other hand, one participant told us he had been raised to be private and cautious about data sharing, and was reluctant to support patient data sharing.

"I'm not comfortable with my data being shared [...] I've been brought up under doctor confidentiality with your patient [...] I think tomorrow everything is going to be done with technology going even further than what it is now. I am just against that." Male, 36, Pakistani

Participants' experiences of health conditions also affected their views on the topic. Those who had experienced more stigmatised conditions were more cautious about the potential of patient data sharing, since it held more potential for discrimination. One participant shared that they had been affected by cancer, and therefore wanted their relevant patient data to be used to improve and develop new treatments. Consequently, they were more in favour of the potential of patient data sharing than some of the other participants.

"I've been on a follow-up programme which is extremely thorough. I don't feel that I'm doing anything that's putting me at risk in any way [...] I'd be happy to give whatever information is wanted by organisations connected with health as long as it's not subject to abuse or misuse..." Male, 69, White British

The impact of communication and transparency

Participants wanted clear guidance on the use of their patient data: why it was used, who by, which specific data was used and how it was kept secure.

Participants felt they should have been informed about the changes brought about by the COPI notices. Again, the topic of GDPR came up during this discussion. The fact that there was little notice and little information on the change negatively affected participants' trust in the scheme, as well as trust in the management of their patient data in general. The lack of communication around this initiative made them feel like their data may be shared in the future with little communication.

Choice

Participants wanted control over how their patient data was shared and used, whether this was through getting the GP to keep certain details of their health and care 'off record' in case it would be picked up for research purposes, or through giving their consent for specific medical research.



“I believe that it should be able to use information, but with everything, it should be your own choice, you shouldn't be tricked into mistakenly having your information out there if you don't want it to be there” Male, 24, Black African

They expressed a strong desire to receive clear and specific information on how their patient data may be used so that each person could make their own choices. This was a key condition even if their patient data was used by an organisation they trusted for a use they supported.

Exploring options for the future

After an open discussion, we provided participants with four suggested options to choose from for the future use of patient data in research. We also asked participants to choose their favourite and least favourite options a few days after the group, to allow them time for reflection and to record their individual perspectives. These options were:

1. Go back to the way things were before the pandemic and keep the data that was collected.
2. Go back to the way things were before the pandemic and delete the data that was collected once the pandemic is over.
3. The rules should be changed permanently to allow some people/organisations to use confidential information without consent for
 - a. Covid-19 related purposes only
 - b. Any health-related purpose
4. The rules should be changed permanently to allow all groups use confidential information without consent for
 - a. Covid-19 related purposes only
 - b. Any health-related purpose

Participants recognised the importance of using patient data to tackle the Covid-19 pandemic. However, they were concerned about the change in the rules being kept forever due to their mistrust in data management in general, which was partly caused by the breaking of the GDPR story.

- **Option 1** was preferred by all participants in the focus group (although this changed in the post group questionnaire) as they wanted to feel in control over how their patient data was used, and have the possibility to consent to its uses. Since the data gathered throughout the Covid-19 pandemic had proved so valuable, participants wanted to ensure the body of knowledge was kept.
- **Option 2** was less popular than the first option, and was perceived negatively since participants felt it would be wasteful to delete the information that had been gathered throughout the pandemic.

Within these options, participants made little differentiation between uses for any health-related purpose or Covid-19-related purposes only, as they saw these options being applied in the long-term and after the current pandemic. Concern was also raised around these options, as participants feared that the new guidelines would be kept and 'quietly forgotten about', posing risks to their data privacy and freedom of choice.

- **Option 3** raised concerns around which groups could be selected to have access to data. Participants wondered how those groups would be selected, and how they would be assessed as being the right groups.
- **Option 4** was the least popular, since participants did not trust commercial organisations to have access to their data without their consent. Some worried that this option was final and would not allow for variation in the future since those organisations would have access to their data in the long term, especially if it were allowed for any health purpose.

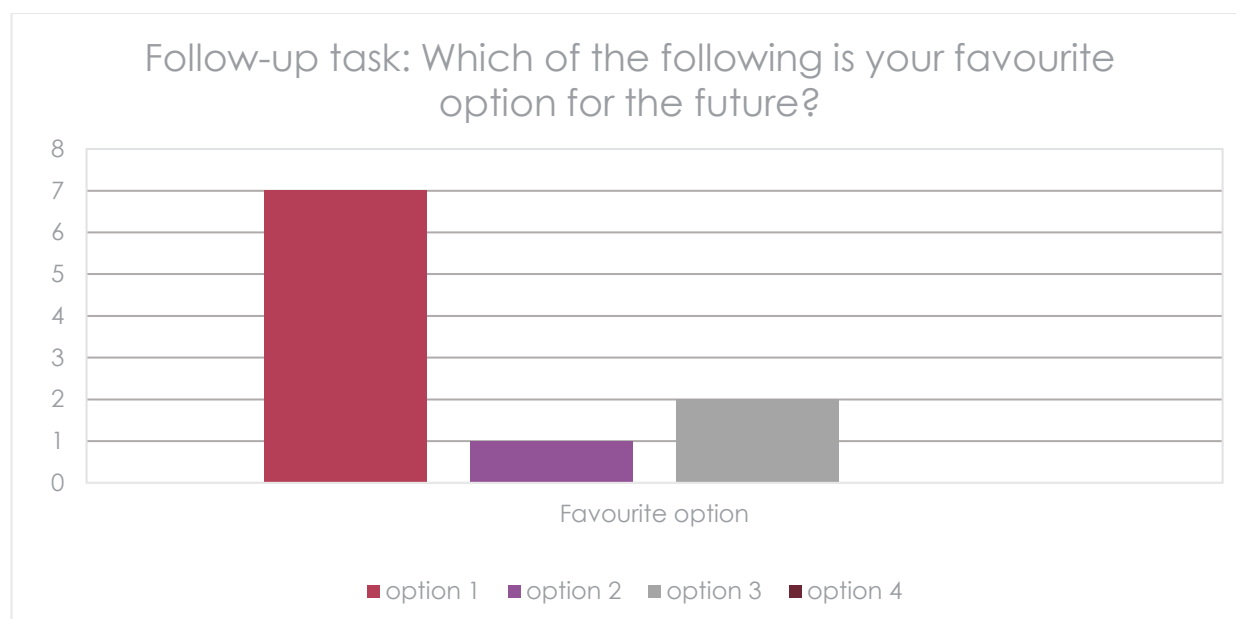
Following the group, we saw some variation in participants' preference. In the group, option 1 was chosen unanimously as the preferred option. However, with a few days' reflection, participants also saw benefits to option 2 and 3, particularly if some of the conditions discussed above were in place.

Participants views were potentially clouded by the more recent conversation around the risks of sharing patient data than the conversation about benefits. There is a strong possibility that the convergence towards option 1 in the context of the group was caused at least partly by groupthink, especially with this research engaging one small group only.

More reassurance may have been needed around options 3 and 4, as participants raised concerns and struggled to see their advantages in group discussion. For instance, they identified no real link between option 3 and the advancements which helped counter Covid-19. It is likely that participants who chose option 3 in the follow-on task reflected on some of these points, or felt more comfortable expressing their preference for group 3 outside of the group.

However, it is important to note that participants expressed a strong desire for mechanisms to control the use of their patient data throughout the session, with consent being a key factor.

"I'm comfortable for my patient data to be made available to many organisations, anonymously, even to pharmaceutical companies. I would not wish data that included my name, address, or any identifying information to be provided other than to the NHS" (Option 3b) Male, 69, White British



Conclusion

Despite concerns with many aspects of patient data sharing, in discussion, the majority of participants were favourable towards its use. They saw the use of patient data in medical research as essential for delivering medical treatment, improving our understanding of health, supporting medical advancement and planning for the future.

Although participants were generally supportive of patient data sharing before the session, the general level of support increased, with more participants strongly supporting the use of patient data in medical research by the end of the focus group.

However, they did raise some key concerns, and wanted to see specific conditions in place to ensure their data was used appropriately. These surrounded:

- **How their data was kept:** they wanted assurances that their data was kept secure from data breaches, leaks or losses.
- **Who had access to their data:** levels of trust varied across different organisations, with participants expressing widespread trust in the NHS. Trust in the government varied across participants, and trust in commercial organisations was particularly low.
- **Why their data was being used:** participants wanted their patient data to be used for health-related purposes, and to be of public benefit. They were concerned around function creep or change of purpose.
- **What data was being shared:** participants wanted their data to be shared in a non-identifying and anonymous way, ideally as part of a large dataset. They also wanted to be certain that the data being shared was proportionate to the research being conducted, and that only essential information was shared.



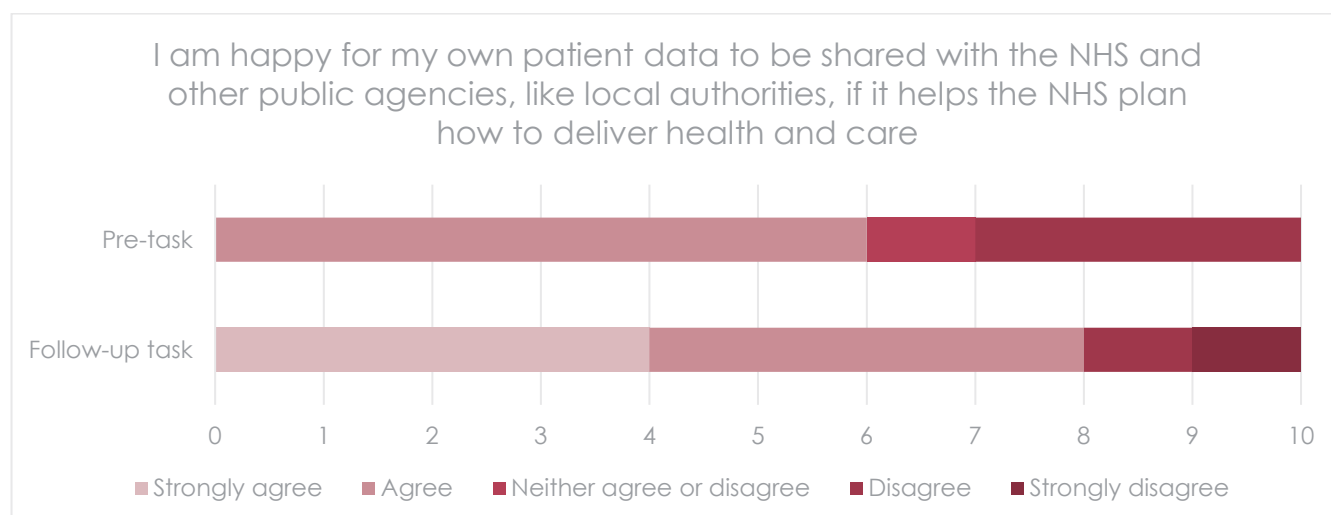
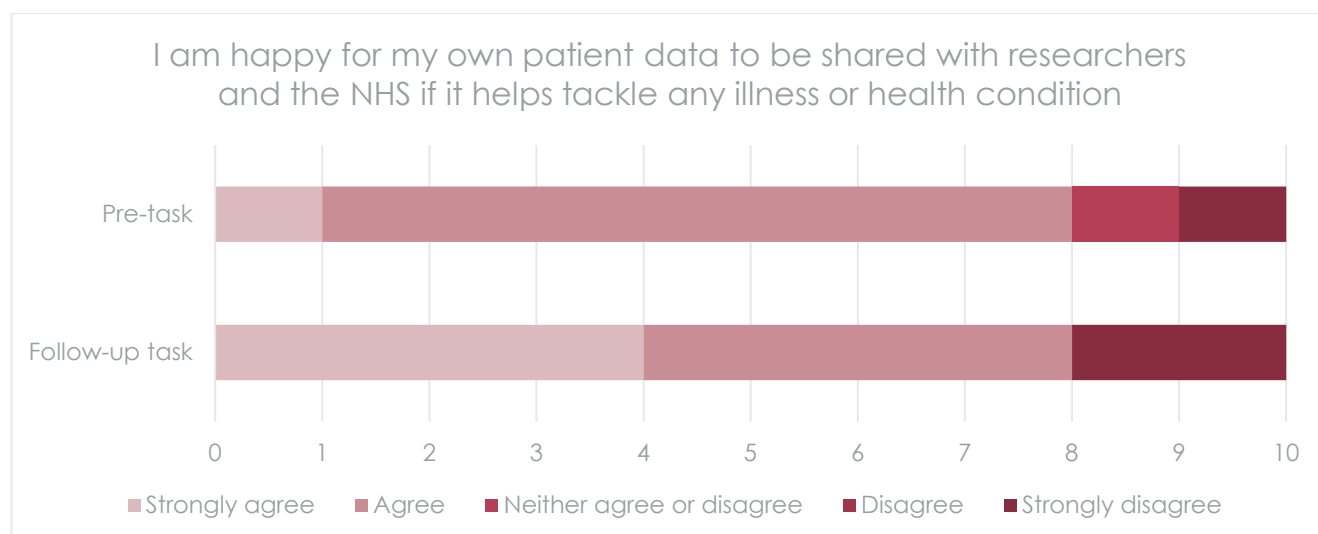
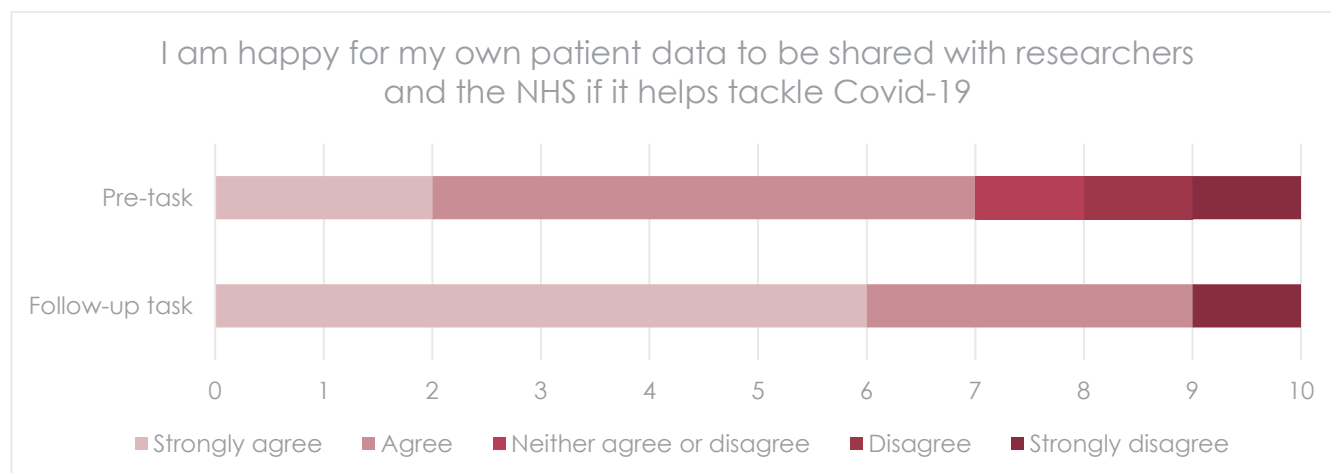
Going forwards, participants identified key conditions to increase their trust in the process of data sharing:

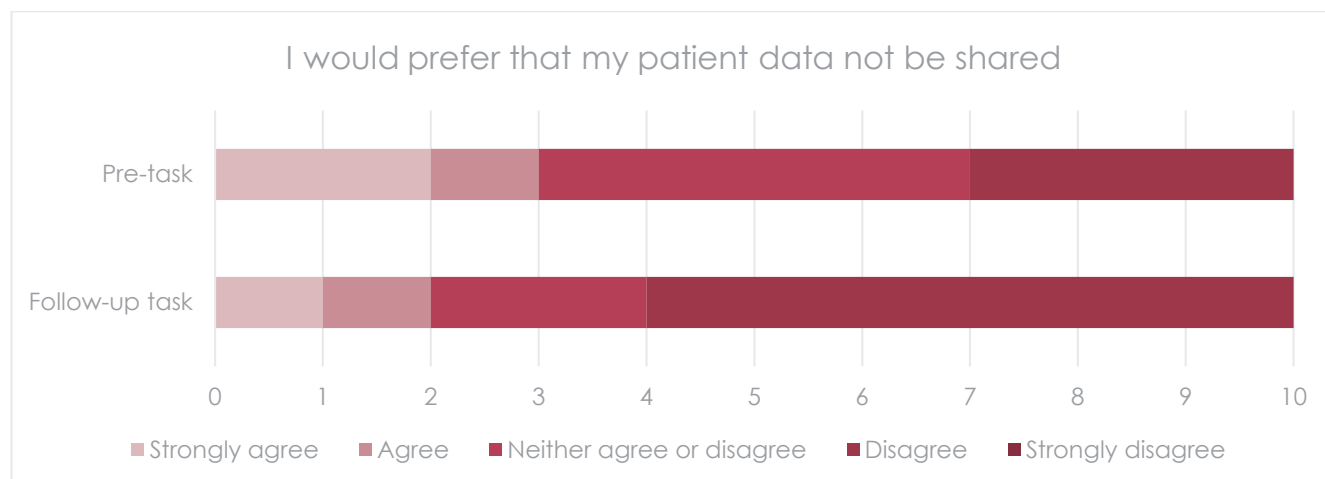
- Participants wanted clear and transparent information about how their data was being used, what 'anonymisation' means, and how this might impact them.
- Participants wanted to feel like they had a choice over how their patient data could be used – ideally by consenting for its use in individual medical research projects.

Appendix

1. Questionnaire responses

The ten participants responded to the pre-task in the days running up to the focus group. The follow-up task was shared with participants the day following the group, and allowed participants up to one week to respond.





2. Questionnaires

1.1.1. Pre-focus group information and questions

Patient data is information about an individual patient. It can include lots of things, like:

- Information about our past and current illnesses and diagnoses
- Information on our past, current and future health and lifestyle, for example if we smoke or drink
- Information about our past prescriptions
- Information on our genetic (family history) or biometric data (fingerprints, eye scans)

Patient data is usually stored by our GP and the NHS to help deliver our care. Patient data can also be used for research: to understand how to develop treatments, monitor population health, monitor the safety of medication and learn more about disease and illnesses.

To what extent do you agree with the following statements? – Likert 1/5.

- I am happy for my own patient data to be shared with researchers and the NHS if it helps tackle Covid-19
- I am happy for my own patient data to be shared with researchers and the NHS if it helps tackle any illness or health condition
- I am happy for my own patient data to be shared with the NHS and other public agencies, like local authorities, if it helps the NHS plan how to deliver health and care
- I would prefer that my patient data were not shared

What do you think the benefits of sharing patient data might be?

What do you think the risks of sharing patient data might be?



1.1.2. Follow up questions

To what extent do you agree with the following statements? – Likert 1/5.

- I am happy for my own patient data to be shared with researchers and the NHS if it helps tackle Covid-19
- I am happy for my own patient data to be shared with researchers and the NHS if it helps tackle any illness or health condition
- I am happy for my own patient data to be shared with the NHS and other public agencies, like local authorities, if it helps the NHS plan how to deliver health and care
- I would prefer that my patient data were not shared

What do you think the benefits of sharing patient data might be?

What do you think the risks of sharing patient data might be?

Please select your favourite of the following options for patient data use after Covid-19 (these are the ones we discussed in the group):

- Go back to the way things were before the pandemic and keep the data that was collected.
- Go back to the way things were before the pandemic and delete the data that was collected once the pandemic is over.
- The rules should be changed permanently to allow some people/organisations to use confidential information without consent for
 - Covid-19 related purposes only
 - Any health-related purpose
- The rules should be changed permanently to allow all groups use confidential information without consent for
 - Covid-19 related purposes only
 - Any health-related purpose

Please select your least favourite of the following options for patient data use after Covid-19 (these are the ones we discussed in the group):

- Go back to the way things were before the pandemic and keep the data that was collected.
- Go back to the way things were before the pandemic and delete the data that was collected once the pandemic is over.
- The rules should be changed permanently to allow some people/organisations to use confidential information without consent for
 - Covid-19 related purposes only
 - Any health-related purpose



- The rules should be changed permanently to allow all groups use confidential information without consent for
 - Covid-19 related purposes only
 - Any health-related purpose

Please explain your answer.

3. Session Sides



Patient data use in research

Focus group



Patient Data- Focus Group



Welcome!



Zoe



Hannah

- **Traverse** – an independent research organisation
- **PHG Foundation** – a think tank working on health policy

2



Patient Data- Focus Group



Why we're here

- We want to understand people's views on how patient data should be used.
- The group will last 2 hours and we'll take a break halfway through.
- We're a small group – we don't all need to agree on things!
 - It's okay to express different opinions: we want to hear all views so feel free to share.
 - There are no right or wrong answers!

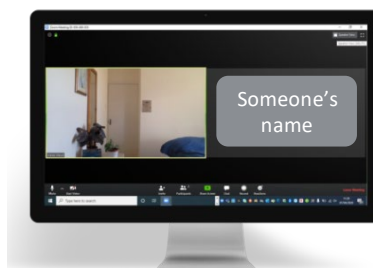
3



Patient Data- Focus Group



Zoom etiquette



- Don't worry about children, pets, or backgrounds!
- Tell us if there's a problem
- Use the chat function for comments and questions
- Keep muted when not speaking
- Raise hands
- Allow everyone time to speak

4





How patient data was used before Covid-19

- Your medical information is private and confidential. Confidentiality is what you expect when sharing information you reasonably expect to be kept private: i.e. between a patient and a doctor.
- Information you share with your doctor must be kept confidential and cannot be shared unless there are good reasons:
 - If it is needed for your care: to schedule hospital appointments, deciding what tests you might need, to record summary information on your health and care record
 - For admin purposes: to pay for the care you received from the NHS
 - In an emergency
 - To prevent harm

7



How patient data was used in research

- For all other reasons, your consent was needed to share private and confidential information.
- This meant that for most medical research, researchers had to ask for your consent before using your patient data, or approval from an official committee to use it without your consent.

8



Discussion



- Your name, where you are speaking to us from and what you last had to eat.
- Reflect on the questions you answered.
- Share your initial thoughts on the topic.

5



What is patient data?

- Patient data is information about an individual patient. It can include lots of things, like:
 - Information about your past and current illnesses and diagnoses
 - Information on your past, current and future health and lifestyle, for example if you smoke or drink
 - Information about your past prescriptions
 - Information on your genetic (family history) or biometric data (fingerprints, eye scans)
- Patient data is usually stored by your GP and the NHS to help deliver your care
- Patient data can also be used for research: to understand how to develop treatments, monitor population health, monitor the safety of medication and learn more about disease and illnesses.

6



How patient data is used in research now

- With the Covid-19 pandemic, the UK government decided to allow patient data to be used for Covid-19 related medical research without the normal limits in place.
- This means that your consent isn't needed, and researchers don't have to ask for specific approval to use your data.
- These rules have been extended to September 2021 and will likely be extended further.

9



Example 1 – GenOMICC study

This study explores the role of genetics in how sick people became with Covid-19. It compares the DNA of people who were severely affected to those of people who were only mildly affected to understand why some people had life threatening symptoms.



Type of data that was used: Health records, genetic data from the person and the virus



Patients whose data was used: Volunteers



How the data was gathered: Blood samples



Researchers who used the data: University of Edinburgh, Genomics England, GenOMICC

10



Example 2 – STOP Covid-19 study

This study is looking to understand the structure of the Covid-19 virus to understand how it changes as it spreads across the globe. This is to help researchers predict the future spread of the virus, estimate the number of worldwide cases and help develop models to estimate the end point of the pandemic.



Type of data that was used:
Genetic data from the virus



Patients whose data was used:
Patients in a hospital in Wessex



How the data was gathered:
Through Covid-19 test swabs



Researchers who used the data:
University of Portsmouth

11



Example 3 – Enhanced Covid-19 testing

In areas where variants that are known to spread more easily have been discovered, enhanced testing has been introduced. In Lambeth, when a case of the South African variant was discovered in February, surge testing was used and positive cases were sequenced for genomic data to help understand Covid-19 variants and track the spread in the area.



Type of data that was used:
Genetic data from the virus



Patients whose data was used:
People living in Lambeth



How the data was gathered:
Covid-19 test swabs



Researchers who used the data:
Local authorities

12



Example 4 – Shielding patients

Local authorities and NHS service providers are able to identify clinically extremely vulnerable people to provide targeted support before, during, and following local outbreaks.



Type of data that was used:
Health records, NHS datasets



Patients whose data was used:
Patients who fit the criteria



How the data was gathered:
Analysis of different datasets



Researchers who used the data:
NHS, local authorities

13



Discussion



14



Use-cases – recap stimulus

GenOMICC study

Comparing the DNA of volunteers to understand what causes some people to be very ill with Covid-19 or not.

Ran by Genomics England and a university.

STOP Covid-19 study

Using genetic data from the virus to understand how it evolves as it spreads around the planet. The virus' data is collected from patients in hospital.

Ran by a university.

Enhanced Covid-19 testing

Local authorities used enhanced testing to analyse genetic data from Covid-19 to identify where variants may be. This data was gathered by Covid-19 tests.

Shielding list

Clinically vulnerable people were identified using NHS datasets and records so that local authorities could provide them with extra support during the pandemic.

15



Benefits and harms

For individuals	Benefits:	- People may be identified and discriminated against - People may lose autonomy / control over decisions which are important to them - Lack of transparency - Others may benefit from information about patients but the patients won't (i.e. companies making money)
For society		- Loss of trust between patients and health professionals - People may stop disclosing information that is important for their care - People might be worried about data sharing within the government (i.e. with DWP) - People may put off going to the doctor

16



Break – 5 minutes



17



Discussion: what should happen next?



18



Potential options for the future

1 Go back to the way things were before the pandemic and keep the data that was collected

2 Go back to the way things were before the pandemic and delete the data that was collected once the pandemic is over

3 The rules should be changed permanently to allow **some groups** to use confidential information without consent

For some purposes

For any health-related purpose

4 The rules should be changed permanently to allow **all groups** to use confidential information without consent

For some purposes

For any health-related purpose

19



What access should organisations have after Covid -19?



Organisations directly involved in your care (GP practices and hospitals)



Public health organisations (Local authorities, Department of Health and Social Care, Public Health England)



Commercial companies

20



Thank you!

- Any questions?
- Last minute reflections?
- Any burning issues or concerns you'd like to raise?

If you want to learn more about the topic, you can visit:
<https://www.phgfoundation.org/>

21



The PHG Foundation is a non-profit think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare and deliver improvements in health for patients and citizens.

intelligence@phgfoundation.org



UNIVERSITY OF
CAMBRIDGE

phg
foundation
making science
work for health