



SNAPSHOT REPORT 1

DATA-DRIVEN PUBLIC POLICY



Rachel Allsopp (Lead Author)
Claire Bessant, Sneha Dawda, Charlotte Emmett, Matthew Higgs, Ardi Janjeva, Marion Oswald
Guangquan Li, Selina Sutton and Mark Warner

With special thanks to our reviewers Alexander Babuta, Keith Ditcham and James Sullivan

18 May 2021

EXECUTIVE SUMMARY

- ‘Data-driven’ decision-making has been at the heart of the response to Covid-19 in the UK. Data-driven approaches include: sharing, linkage and analysis of different datasets from various sources; predictive modelling to anticipate and understand transmission and inform policy; and data-driven profiling to identify and support vulnerable individuals. This Snapshot Report incorporates OMDDAC’s findings from interviews with stakeholders, together with published research, to capture the lessons learned throughout the pandemic across these three case studies.
- Several key themes have been identified across the three case studies. These include issues of transparency with regard to the role of data in the public policy decision-making process; concerns around poor data quality (including errors, omissions and outdated data) and a lack of interoperability across public sector datasets; and the identification of skills gaps in relation to data literacy and information governance.
- The importance of robust data protection safeguards becomes more pronounced as data sharing increases. Information governance should therefore be ‘front-loaded’, to ensure that it forms an integral part of any data acquisition process from the outset. However, as has been demonstrated during the pandemic, this need not represent a barrier to data sharing.

HEADLINE RECOMMENDATIONS

- **Data standards** - Government investment in public data architecture is urgently required to address data quality issues and improve interoperability across public sector data (including both health and non-health bodies). A data standards framework should also be agreed and implemented across health and non-health bodies to ensure improved standards are maintained.
- **Skills gaps** - Central and local government resource and training should be directed to 2 areas:
 - (1) **data literacy**: to bridge the gap in understanding between mathematical modellers and decision-makers and ensure decision-makers are well-versed in both the capabilities and the limitations of quantitative data. This should be supplemented with contextual information in the form of qualitative data where appropriate (for example, to incorporate localised knowledge);
 - (2) **information governance**: to avoid unnecessarily risk-averse attitudes towards appropriate data-sharing by information governance professionals across the public sector, where the necessary safeguards are in place.
- **Transparency** - Central and local government should, where possible, make data publicly available (together with its interpretation) that has been used in reaching policy decisions. An explanatory notice should also accompany publication of policy decisions, which makes clear the additional factors and values taken into account, and highlights any limitations, uncertainties, and assumptions. Efforts should also be made to directly engage the public through consultations and education campaigns on the role their data can play in public policy and practice post-pandemic.

CASE STUDY 1

DATA SHARING, LINKAGE & 'INNOVATION'

HEADLINE FINDINGS

- The **sharing and linkage of data** from a diverse range of sources has been central in driving the national Covid-19 response, both centrally and locally.
- The pandemic has demonstrated that an open approach to data sharing can lead to **better informed decision-making** and **greater public good**. Stakeholders are therefore keen to build on this progress though, at the same time, stressed the need to ensure **robust frameworks for information governance** are in place: 'good governance becomes more important now that more information has been shared'. Provided this is the case, data protection should not be perceived as a barrier to appropriate data sharing.
- Greater **transparency** is required regarding the role of data as part of decision-making processes, including its **limitations**, to ensure **proper public scrutiny**. This should be addressed by, where possible, making the data (together with its interpretation) available publicly. Decisions should also be accompanied by explanatory justifications which identify the factors considered and highlight any limitations within the data. Decision-makers should also be mindful that 'some things cannot be mathematically reduced to facts' and therefore seek to strengthen quantitative datasets with qualitative insight, for example using local knowledge to provide additional context.

CASE STUDY OVERVIEW

During the pandemic, public bodies (including national and local government departments, the NHS, the police,¹ and research bodies such as the ONS) combined existing and novel datasets in unprecedented and 'innovative' ways to better understand the prevalence of Covid-19. This informed public policy decisions aiming to prevent and contain outbreaks of the virus. The types of data being shared are diverse, encompassing both health² and non-health³ datasets. Moreover, the data is not only sourced from the public sector. It is also provided by a range of private sector sources (e.g. Google, social media platform providers, retail, and telecoms providers). One especially novel data source which has been incorporated into this '**data ecosystem**' is wastewater. The Government's wastewater monitoring programme analyses samples to detect and quantify the amount of viral RNA load from

'The intriguing thing about wastewater is that it knows no statistical boundaries – the wastewater system reflects the way people actually live'.
Stakeholder interview G10

¹ We explore the sharing of data with police in detail in our third report of this Snapshot series: 'Policing and Public Safety')

² E.g. GP, hospital and test & trace data.

³ E.g. employee and school absence data and mobility data, such as telecoms, transport, and geo-location data.

Sars-CoV-2,⁴ to help provide an early warning system for new cases in a specific geographical area.⁵ This can then be used in combination with other data to follow up with local measures such as community testing or targeted messaging to contain outbreaks.

AIMS, OBJECTIVES & BENEFITS

Data has been central in driving the national Covid-19 response in the UK. It has been used to determine the Covid-19 alert level,⁶ and inform national decisions on lockdown restrictions.⁷ The sharing of data from central government to local authorities, particularly NHS Test and Trace datasets, has also been crucial in enabling a localised pandemic response. Examples include deciding where to locate testing and vaccination centres, ‘knocking on doors’ and instructing individuals to self-isolate. More generally, the sharing and linkage of data is an essential primary step for the production of robust statistical models, the outcomes of which have been central to pandemic decision-making (as we explore in Case Study 2).

‘We’ve learned the importance of mobility data for helping with the pandemic response all the way across the board from modelling ‘R’, to understanding how to design lockdowns, what the parameters of those lockdowns are, the geographical areas for those locales... There is a sense of innovation there which is understanding the datasets that are designed for one purpose actually can give you insights in a whole different area... [which has been] really important in understanding non-observed pandemic spread’.

Stakeholder interview G9/10

The increased data sharing was ‘set up in recognition of the holes and gaps, exposed by the pandemic, in data coverage [as well as the] ability to analyse [and] manipulate those data and provide insights to actually create an effective response to the pandemic’.⁸ By ‘bridging the gap’ between health and non-health data, it has been argued that ‘linking the data can provide value-added insights to inform policy’, leading to ‘better public decision-making informed by the facts’.⁹

Many stakeholders felt that the pandemic has created a greater awareness of the need for a more open approach to data sharing and linkage. This has helped to ‘generate even better and more quantitative evidence’, that can ‘lead to better decision-making and therefore greater public good, which could include in this particular discussion saving lives’.¹⁰ Moving forward, stakeholders considered that ‘the pandemic [has] assured us really clearly what we can do, what we can achieve with data ... but it also showed us really clearly what we can do better... to drive strategic improvements in our data ecosystem and infrastructure [so that] the position is better for the future’.¹¹

⁴ The ‘severe acute respiratory syndrome-related coronavirus 2’ termed coronavirus disease 2019 (Covid-19). SARS-CoV-2 is an enveloped, positive-sense, single-stranded RNA virus that, upon infection of a host cell, deploys a ‘translation-ready’ RNA molecule, which uses the protein synthesis machinery of the host to express a set of viral proteins crucial for replication: Tay, M.Z., Poh, C.M., Rénia, L. et al. The trinity of Covid-19: immunity, inflammation and intervention. *Nat Rev Immunol* 20, 363–374 (2020). 8

⁵ <https://deframedia.blog.gov.uk/2021/04/26/tracking-covid-19-in-wastewater/> and <https://www.theguardian.com/world/2021/apr/24/were-the-poo-crew-sleuths-test-for-covid-by-reading-signs-in-sewage> accessed 29 April 2021

⁶ <https://www.gov.uk/government/publications/uk-covid-19-alert-level-methodology-an-overview/uk-covid-19-alert-level-methodology-an-overview> accessed 3 May 2021

⁷ <https://www.gov.uk/government/publications/covid-19-response-spring-2021/covid-19-response-spring-2021> accessed 3 May 2021

⁸ Interview with Government representative (‘G’)9/10

⁹ Interview with G9/10

¹⁰ Interview with Data expert (‘D’)10

¹¹ Interview with G17

RISKS & CHALLENGES

Data limitations: Stakeholders reported that the quality of the data was variable, often containing errors, omissions, and outdated information. Moreover, data interoperability was reported to be an issue due to inconsistencies in data labels and data ontology: 'bringing [the data] together requires lots more engineering and familiarisation... than would be necessary if we had widespread data standards'.¹² It was thus recommended that a data standards framework¹³ be developed and implemented to address these operational issues. More fundamentally, stakeholders pointed out that 'some things we are observing during the pandemic cannot be mathematically reduced into facts and requires some form of interpretation, coupled with local knowledge'.¹⁴ This qualitative, contextual element is missing in the employment of data-driven approaches drawing primarily on quantitative, administrative datasets, which can significantly impact upon understanding. Several local authorities have therefore sought to supplement administrative data with findings from qualitative research, employing surveys, interviews, and local knowledge more generally, to provide additional context. 'Often the quantitative data will tell you what is wrong and we are reliant on the qualitative information to tell us why'.¹⁵ Indeed, local knowledge has proved critical in informing more targeted responses to the pandemic: '[we have] local knowledge of individual neighbourhoods, quirks, preferences of how they want to be communicated [with]. Devolving to a local level [has been] so important'.¹⁶

'At the policy level if you're making choices based on data that affect people's freedom of movement and life in general, you really need to go "to the max" to ensure the independence and quality of the data, and you also need to assert on a wider national interest basis what the policy responses are.'

Stakeholder interview G3

The role of value judgements and assumptions:

Relatedly, several stakeholders were concerned that in emphasising the importance of the data, it gives the impression that a decision is wholly objective and 'beyond dispute'.¹⁷ However, as one stakeholder summarised: 'the actual situation is far more complex. Those decisions are in turn informed by value judgements on which datasets are actually being used

and whose data is actually being evaluated... We then need to make value judgements about what the data implies and what we are going to do with it. If we understand that facts are theory laden and value laden, then any use of the data to inform decisions implies values'.¹⁸ This issue is further compounded when taking into account the application of statistical modelling with its associated levels of uncertainty (see Case Study 2). Stakeholders also highlighted that many decisions relating to the pandemic are 'essentially political', involving 'quite properly' the weighing up of risks and other factors: for example, 'opening schools or not relaxing lockdowns... those decisions involve other aspects like care...negative impact on mental health and economic impact. These aspects are hard to compute, so the decision is complicated'¹⁹.

¹² Interview with G17

¹³ An agreed set of principles for managing the quality of data which is applied consistently across all bodies, for example, sharing data across the public sector. See for example the Government Data Quality Framework (December 2020)

<https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework>

¹⁴ Interview with Regulatory body representative ('R')1

¹⁵ Interview with G15/16

¹⁶ Interview with G15/16

¹⁷ Interview with Third sector representative ('T')1

¹⁸ Interview with Academic ('A')2

¹⁹ Stakeholder interviews – Data expert ('D')6; G9/10; A8 (respectively).

Privacy: Stakeholders acknowledged that ‘there will inevitably be some people in the general public who think that too much data is being shared’²⁰ and, relatedly, will be concerned about the potential for ‘mission creep’ with data being used as a tool for more general surveillance.²¹ As will be expanded upon in the following section, however, much of the analysis, especially at population-level, is carried out on aggregated data to identify patterns and trends: ‘we’re not surveilling people, we’re not using individual data’. Notwithstanding the initial data collection, several stakeholders felt that privacy concerns were minimised in this way of working. Indeed, several stakeholders were keen to point out innovations in setting up privacy-preserving architecture for sharing data responsibly, such as OpenSafely,²² as well as strong information governance frameworks, which ‘would not have happened but for the pandemic’.²³

Transparency: In response to these issues, stakeholders called for greater transparency regarding how decisions are reached, including the data itself and its limitations, as well as the additional values which form part of the decision-making criteria. Whilst there was acceptance that it is more difficult to be transparent when ‘moving at pace’ in an emergency context (which relatedly highlights the need for pre-pandemic planning in this regard), stakeholders felt that this issue was of paramount importance given the unprecedented restrictions on liberty that have been introduced. As the Public Administration and Constitutional Affairs Committee stated:

*‘there is a moral imperative on Government to clearly justify each of their decisions. Part of that is making the data that is driving the response, and its interpretation, available so people can understand why they are being asked to make such sacrifice... The Government must build trust and co-operation by being open and transparent about the data. Data transparency is not just a moral issue, it is integral to the success of the response to this pandemic. Transparency builds trust, and trust aids compliance with rules’.*²⁴

LAW, REGULATION & GOVERNANCE

Legal frameworks

In relation to the sharing of health data particularly, the common law duty of confidence has been overridden temporarily during the pandemic by the Secretary of State through a series of ‘COPI notices’.²⁵ This has allowed the sharing of confidential information

Duty of confidence:
A common law duty of confidence arises whenever information of a personal, private, or intimate nature is disclosed in circumstances where it is reasonable to expect that it would be kept private. Information shared with health and social care professionals will often give rise to a duty of confidence (*AG v Guardian Newspapers (No.2)* [1990] 1 AC 109). This common law duty is reflected in professional codes of ethical guidance and NHS contracts of employment. The duty is not absolute: implied consent is sufficient to justify information sharing for the primary purpose of patient care or for local clinical auditing. Otherwise, a patient’s identifiable information cannot normally be disclosed unless there is: explicit consent; a legal requirement for data disclosure; or a clear overriding public interest to disclose which outweighs the obligation of confidence.

²⁰ Interview with D10

²¹ Or being shared for the additional purposes of police enforcement. See Snapshot Report 3: Policing and Public Safety for further consideration of this issue.

²² <https://opensafely.org/> accessed 3 May 2021.

²³ Interview with T1

²⁴ Public Administration and Constitutional Affairs Committee, *Government transparency and accountability during Covid 19: The data underpinning decisions* (HC 2021 803) <https://committees.parliament.uk/publications/5076/documents/50285/default/> accessed 3 May 2021.

²⁵ NHS Act 2006, s.251; and the Health Service (Control of Patient Information) Regulations 2002 (COPI)

without consent for Covid-19 purposes only.²⁶ These 'COPI notices', which are currently effective until 30 September 2021, require NHS Digital, NHS England & Improvement, all healthcare organisations, arms-length bodies, local authorities and GPs (including a specific requirement related to the UK Biobank project) to process and share confidential patient information with organisations involved in the response to coronavirus for Covid-19- specific public health, surveillance and research purposes.

Data Protection:

The Data Protection Act 2018 and UK GDPR apply where an organisation processes (collects, stores, shares etc) personal data (information that relates to an identified or identifiable living person – Art. 4). The legislation will apply where data may be used to identify an individual directly or in combination with other information, using all reasonable means possible. Examples of identifiers include: name, location data, online identifiers such as IP address. The use of 'special category data' (including data concerning health, race or ethnic origin) imposes additional, more stringent requirements (Art. 9).

Much of the data being used by the Government to inform public policy (aggregated datasets of anonymised data) will fall out of scope of the data protection legislation. However, there will be datasets which may not provide details of individual names yet still risk deanonymisation if the parameters of the dataset are sufficiently narrow. For example, where wastewater is used to monitor levels of the virus, stakeholders discussed the challenges in determining whether they were dealing with personal data, colourfully sharing that 'it depends how far up the pipe you go'.²⁷

Clearly test and trace data, designed to identify individuals who should self-isolate, will fall in scope. Where the data is (potentially or actually) identifiable, any sharing must comply with the seven core data protection principles.²⁸ A lawful basis for processing must be identified, for example: consent; where processing is necessary to protect the vital interests of the data

subject or another person; or to perform a task carried out in the public interest or in the exercise of official authority.²⁹ For special category data any consent must be explicit, though where explicit consent cannot be obtained, alternative justifications may include: processing necessary for reasons of substantial public interest which is proportionate to the aim pursued and affords suitable, specific measures to safeguard the data subjects rights; processing necessary for reasons of public interest in the area of public health and again on the basis of suitable and specific measures to safeguard the data subjects rights including rights to confidentiality.³⁰ Similarly in respect of the right to privacy more broadly, certain interferences must be proportionate. Applying this to the analysis of wastewater, whilst this may result in the identification of a household (and thus potentially an individual) as being infected with Covid-19, the disclosure of such information to PHE may nonetheless be justified for the protection of health provided no more information is provided than necessary and only to a limited number of individuals.

²⁶ These include: disease surveillance for the purposes of research, protecting public health, providing healthcare services to the public and monitoring and managing the Covid-19 exposure. Only the minimum amount of data necessary to inform the Covid-19 purpose shall be shared and information anonymised where it is practicable to do so (Reg 7).

²⁷ Interview with G9/10

²⁸ The seven core data protection principles are enshrined in UK GDPR, Art. 5 as follows: Processing must (1) be lawful, fair and transparent (i.e. data should not be used in a way that data subjects would not expect and subjects should be informed of the use of their personal data); (2) data should be used for limited purposes identified at the outset of processing or for other compatible purposes; (3) data should be sufficient for the purpose, relevant to that purpose and no more data should be used than necessary; (4) data should be accurate; (5) data should be stored for no longer than necessary; (6) data should be stored securely; and (7) processes should to be in place to demonstrate UK GDPR compliance.

²⁹ UK GDPR, Art. 6.

³⁰ UK GDPR, Art. 9.

Public sector organisations also have obligations under equality legislation,³¹ meaning that considerations of equality should be incorporated into the day-to-day business of public authorities. This includes deployment of measures to tackle the Covid-19 pandemic and considering how such measures might affect different groups. For instance, a vaccine or testing programme may adversely impact young people with a disability if triaging measures are purely based on age. In addition, public sector services must not directly or indirectly discriminate based on a protected characteristic.³² Centralised technological approaches to testing, booking vaccinations or directions to shield are all examples of pandemic responses which could require justification against equality concerns, should they disadvantage certain groups (for example the elderly or the visually impaired).

Stakeholder experience

Throughout this research, stakeholders recognised the importance of having robust information governance frameworks in place. Indeed, information governance proved to be a particular barrier to data sharing at the beginning of the pandemic, when local authorities experienced significant difficulties in gaining access to the granularity of data required to inform local responses. More broadly, many stakeholders commented on a culture of ‘institutional nervousness about sharing, linking and repurposing of data’³³ wherein implementation of information governance has been unnecessarily ‘risk-

‘We [had] not been out of lockdown since March, but there was no data to support it at that point. There was no data to say where the hotspots were ... We [needed] to know who these people were, where they live, who they live with, where they work, how they travelled to work... We just could not get access to the data. There were lots of blockages, lots of arguments ... and it took forever for them to start giving us some data and it still wasn't enough. They weren't giving us specific houses or people; it was just aggregated data showing geographical hotspots that we would have guessed anyway. They were not giving us the detail, for a time, to do anything with, so that was that was incredibly frustrating’.

Stakeholder interview G6

averse’: ‘data governance teams have not been sufficiently resourced and trained, and at middle management level there is a tendency that as soon as data governance comes up the word ‘no’ is associated with it... As soon as we mention the word ‘postcode’, you can guarantee that the response will be ‘no’... we have a responsibility if we are collecting data, we should be using those data, or we shouldn't be collecting them, but the answer is not to collect data and then lock it away in a safe’.³⁴

Nonetheless, stakeholders have observed that some of this ‘cultural resistance has evaporated in the face of the pandemic’.³⁵ As one stakeholder noted, ‘now that dam seems to have been breached, insofar as people are still doing things lawfully and in accordance with good governance’– indeed ‘good governance becomes more important now that more information has been shared’ – ‘It came down to will from senior figures in key organisations to overcome these barriers, backed up by ICO (Information Commissioner’s Office) guidance on sharing and anonymisation. The pandemic has made it a pressing enough issue for them to have to nail their colours to the mast’.³⁶ A number of stakeholders commented that they were able to re-examine and ‘operationalise’ data protection

³¹ Primarily the Equality Act 2010. See Case Study 3 for further details regarding this legislative framework.

³² Subject to certain exemptions, see Case Study 3

³³ Interview with R4

³⁴ Interview with G9/10

³⁵ Interview with D3

³⁶ Interview with R4

legislation in conjunction with other existing statutory provisions,³⁷ for the purposes required during the pandemic. Stakeholders reflected that ‘it’s not the law that has been the problem; it’s the issue of shared responsibilities... in 99% of situations, data protection law allows you to do what you want to do, provided the safeguards are there. We have seen the value of good information governance’.³⁸ As another observed, by ‘front-loading data governance and making data governance an integral part of the data acquisition process, right from the start, you don’t end up in what I think are almost unnecessary, quite unhelpful dynamics as you go through’.³⁹

The importance of good data practice (and related concerns about a failure at central government level to follow good practice) was raised by Councillor Ian Hudspeth, Chair Local Government Association Community Wellbeing Board in written evidence to the Public Administration and Constitutional Affairs Committee,⁴⁰ where he highlighted the failure to provide accurate, up to date data to local authorities nationally. He noted further that central government too often focus on ‘what restrictions need to be placed on sharing data, rather than thinking about how we can better use data collectively while meeting data protection requirements.’ Looking to the future he stated that ‘improvements have been made to the data flows where government has co-designed the approach with local government, ensuring that councils’ requirements, as the key users of the data, are incorporated into the design and that it is proportionate and fit for purpose. This approach should be used across all the data systems.’

LESSONS LEARNED

- To ensure **transparency**, decision-makers should make the data used (together with its interpretation) available for public scrutiny. Publication of the data should be accompanied by an explanatory justification detailing how decisions are reached, referencing the data used (and its limitations), as well as any additional factors and values which form part of the decision-making process. Relevant qualitative data (for example local knowledge) should also be incorporated into the decision-making process to support and contextualise administrative datasets.
- A **data standards framework** should be agreed across public sector bodies (including both health and non-health bodies) and implemented to address issues around quality and interoperability of public sector data.
- Data protection need not represent a barrier to data sharing and processing provided robust safeguards are in place. **Information governance** becomes more important with the increase in data sharing and should be **‘front-loaded’** so that it forms an integral part of the data acquisition process. Information governance professionals may benefit from **additional resource and training** to preserve the collective benefits realised by the increased data sharing whilst maintaining robust safeguards.

³⁷ Such as the Statistics and Registration Service Act 2007 (as amended by the Digital Economy Act 2017) for official statistics; and in the health and social care context, the Health and Social Care Act 2012 and the Care Act 2014, as discussed further in Case Study 2 below.

³⁸ Interview with R2/3

³⁹ Interview with G9/10

⁴⁰ In evidence given to the Public Administration and Constitutional Affairs Committee inquiry, on 9th November 2020 <https://committees.parliament.uk/publications/3478/documents/33466/default/> accessed 3 May 2021.

CASE STUDY 2 PREDICTIVE PANDEMIC MODELLING

HEADLINE FINDINGS

- Closely linked with the innovation in data linkage outlined in Case Study 1, the role of mathematical modelling has been threaded throughout critical policy decision-making around the pandemic. Outputs of **statistical modelling** are used to **support the decision-making process in combination with human judgement**, considering a range of additional factors and values.
- Greater **transparency** is needed to improve public understanding around **uncertainty** and the role of **assumptions** in statistical modelling to ensure that outcomes are not represented, or perceived, as being the 'truth'. To address this, there should be disclosure of the relevant additional values and judgements, together with an explanation of how they have been incorporated into the final policy decision.
- There appears to be a **skills gap** between modellers and decision-makers, particularly in public sector leadership. This gap in understanding must be addressed to ensure that decision-makers are appropriately equipped to make use of outputs of modelling.

CASE STUDY OVERVIEW

Predictive modelling has been a key component of policy decision-making during the Covid-19 pandemic. Many of the quantitative indicators used to make decisions cannot be observed directly and must be estimated using the data that is available at a certain point in time. This requires expert knowledge to design models representing the relationship between the data that can be directly measured, and indirectly observed quantitative indicators that need to be estimated for policy decision-making. Over time, the availability of data has improved, and models have evolved. One of the most important quantitative indicators - with which most of the country is now familiar on account of the Government's No.10 Downing Street Covid-19 press conferences - is the R number. This case study will explore the key issues raised by predictive modelling, employing the R number as a specific illustrative example.

The **reproduction number**, also known as the R number, quantifies how many people, on average, an infected person will pass the virus onto. An R value greater than 1 indicates an infected person will, on average, infect more than one other person while an R value below 1 suggests a slowdown of the spread of the virus. This is an important tool to measure and monitor the spread of Covid-19. Over the course of the Covid-19 pandemic, many policies have relied upon the R number, including the roadmap to easing lockdown in the UK.⁴¹ The R number cannot be directly measured but is the combined result of modelling and data. A wide range of data sources, included recorded deaths,

⁴¹ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/975909/S1182_SPI-M-O_Summary_of_modelling_of_easing_roadmap_step_2_restrictions.pdf accessed 3 May 2021.

hospital admissions data, data from contact surveys and household infection surveys, are synthesised, through modelling, to estimate R. In the UK, estimation of R has been carried out by a number of modelling groups based on different sets of assumptions and different data sources. These individual estimates are then combined (at the Science Pandemic Influenza Modelling group (SPI-M) – a subgroup of SAGE) and the R number is presented as a range within which the "true value" is likely to fall.

Processes of estimating 'R' are complex, involving the development and combination of multiple modelling elements with a large set of parameters, the estimation of which relies on a multitude of data sources. The diagram in Knock et al 2021⁴² (below) illustrates the complexity of such modelling. Also shown is the flow of information, occurring both within each individual modelling element and between different modelling elements. This information flow allows two things to happen:

1. **Evidence synthesis**, reconstructing the environments by bringing various constituent parts (different data sources and different model elements) together. R, as well as other relevant quantities, is estimated within such reconstructed, modelled environments, utilising all available data.
2. Second is **propagation uncertainty**. Inside every model are parameters, quantities that are not known to the modellers. They can be estimated through data, but all estimates have uncertainty attached. Through the interconnected nature of modelling, uncertainty in one part of the model can be accounted for in the estimation of the quantity of interest, the R number, in this case. Mathematical/statistical modelling plays a crucial part.⁴³

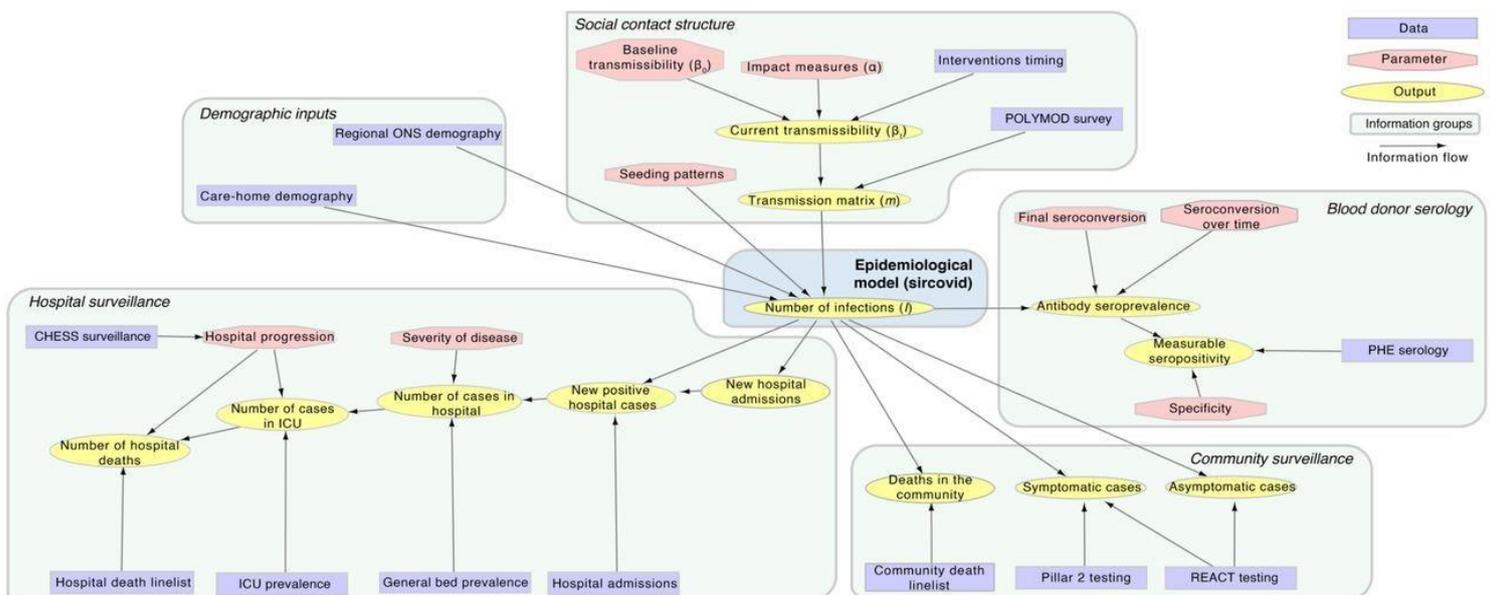


Figure 1 - Knock et al. (2021)⁴³

⁴² Edward S Knock et al, 'The 2020 SARS-CoV-2 epidemic in England: key epidemiological drivers and impact of interventions' (2021) medRxiv 2021.01.11.21249564; doi: <https://doi.org/10.1101/2021.01.11.21249564> (pre-print) accessed 3 May 2021.

⁴³ Knock et al. 2021 (n 23).

AIMS, OBJECTIVES & BENEFITS

Estimates of R can not only be used to indicate the general trend of the pandemic but also used in projecting the potential impacts of each stage of lockdown easing. For example, the SPI-M-O summary of modelling reports the projection of daily hospital admissions in England over a range of R values to examine impacts from the relaxation of restrictions to be introduced on 29th March in England.⁴⁴

More broadly, statistical modelling and data science allow additional value to be extracted from data regarding potential future outcomes, informing key public policy decisions: '[data] gives you a snapshot of the present and the past, but it doesn't give you the future. For that you need models and forecasts'.⁴⁵ According to stakeholders, the role of data more generally, and modelling specifically, has been 'absolutely fundamental' for policymaking. 'At times there have been real policy struggles over lockdowns, whether to keep schools open or closed, the recent scheduling of the unlock, the choice of tiering during the period when we were tier based. The role of mathematical modelling has been threaded throughout that critical decision-making. SPI-M have been incredibly influential throughout Covid'.⁴⁶ Stakeholders also highlighted the potential value in data science and modelling for communication of complex information – 'data science can find simple pictures out of mind-bending data' – though many noted that this potential is perhaps not yet being fully realised.⁴⁷

RISKS & CHALLENGES

Transparency: As explained above, statistical modelling inevitably involves a degree of uncertainty which is addressed using assumptions. 'Assumptions are crucial – all models have assumptions baked into them'.⁴⁸ It was the view of some stakeholders that additional steps could be taken to improve levels of transparency around these assumptions to aid understanding and enable proper scrutiny: '[the government briefings represent] the most prolonged example of regular presentation of data sources to the public. This has been effective in focusing on [the] rise and fall of case numbers. What has been less transparent are the models being used and the assumptions going into the models'.⁴⁹ 'Behavioural parameters', for instance, were identified as being 'incredibly difficult to measure'.⁵⁰ Stakeholders felt that greater transparency was needed to ensure that the outcome of statistical modelling is not presented, or perceived as, the 'truth': 'all models are wrong, but some are useful... it is very easy to confuse the output from your statistical analysis as truth. Mathematicians may not always be good at communicating either explicitly or implicitly the **confidence** we should have in the answer'.⁵¹

⁴⁴ SPI-M-Operational Sub-group, Summary of further modelling of easing restrictions – roadmap step 2 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/975909/S1182_SPI-M-O_Summary_of_modelling_of_easing_roadmap_step_2_restrictions.pdf accessed 3 May 2021.

⁴⁵ Interview with Private sector representative ('P')1

⁴⁶ Interview with D9

⁴⁷ Interview with G3

⁴⁸ Interview with P1

⁴⁹ Interview with T2/3

⁵⁰ Interview with D9

⁵¹ Interview with A3

The power of modelling: Relatedly, stakeholders were concerned that statistics and data science could be used 'beyond its power'.⁵² 'There is an attitude that you throw the data over the wall and the mathematicians throw back the answer, and the answer becomes "the science"'.⁵³ As one stakeholder explained, 'it's modelling for illustration, modelling for building understanding, modelling for storytelling and that is it. Because that's such an important part of the epistemology of mathematical modelling and people often assume a naive "push button, computer tells answer" - this is not what I think the purpose of mathematical modelling is'.⁵⁴ As discussed in Case Study 1, decisions also involve a degree of human, value-based judgment and, again, it was felt that there was a need for greater transparency in this regard.

'I keep telling political leadership not to use maths beyond its power. People often ask you to solve a problem with modelling or statistics. Human judgement also needs to be involved. Maths can only tell you so much ... because we can't measure those behavioural parameters, I've been stubborn in not coming up with a number that is spurious and made up because people tend to believe that a spurious and made-up number that comes from some mathematics and computers is more reliable than the spurious and made-up number that's been invented, and it's not.'

Stakeholder interview D9

Skills gap: In part owing to the complexities outlined above, stakeholders identified a gap in understanding between modellers and decision-makers, particularly in public sector organisations. This 'breakdown at the interface between analysing data and deciding what to do with it'⁵⁵ has led some stakeholders to question whether policy makers, for example, are appropriately equipped to make use of outputs of modelling, leading to concerns around the perception that the Government is hiding behind 'white coats... for its tough decision-making'.⁵⁶ To 'bridge this gap', stakeholders highlighted the importance of meaningful, interdisciplinary interaction between 'specialists in the methodology...and those who are specialists in the domain'.⁵⁷ In addition, the Royal Statistical Society (RSS) has been involved in the establishment of a professional framework for data science,⁵⁸ including a chartered status, to address the 'challenge, particularly in the public sector, that the leadership doesn't have the expertise' and also the lack of 'framework for what they're looking for in terms of skills' when hiring.⁵⁹ More broadly, this complexity has impacted on the ability to effectively communicate the statistics to the public: 'there was a lot of difficulty even conceptualising how to start communicating with the public and that has been exacerbated by quite how difficult some of the things they have had to present are'.⁶⁰ There appears to be a recognition of a growing need for 'data literacy for everybody'.⁶¹

⁵² Interview with D9

⁵³ Interview with A3

⁵⁴ Interview with D9

⁵⁵ Interview with A3

⁵⁶ Interview with G3

⁵⁷ Interview with G9

⁵⁸ <https://rss.org.uk/news-publication/news-publications/2020/general-news/professional-standards-to-be-set-for-data-science/> accessed 3 May 2021.

⁵⁹ Interview with D4

⁶⁰ Interview with D4

⁶¹ UK AI Council, 'AI Roadmap'

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949539/AI_Council_AI_Roadmap.pdf accessed 3 May 2021.

LAW, REGULATION & GOVERNANCE

Legal frameworks

The importance of using personal data for statistical purposes is recognised in the data protection legislative framework.⁶² A body which processes personal data for statistical purposes is exempt from compliance with certain obligations under the UK GDPR, provided safeguards are in place, for example, implementation of technical and organisational measures (such as pseudonymisation). The organisation must also ensure that only as much data as is necessary is used to fulfil the purpose of processing. Article 89 explicitly suggests that where statistical purposes can be achieved by processing of personal data so that it is no longer possible to identify individuals (or where the risk of identification would be remote) such anonymisation should be undertaken. The data protection regime does not apply to anonymised data.

The use of statistics by Government and specifically the production of official statistics is governed by the Statistics and Registration Service Act 2007 and the Code of Practice for Statistics,⁶³ which details practices that producers of official statistics must adhere to when producing and releasing official statistics and emphasises three values: trustworthiness, quality, and value to society.

The UK Statistics Authority has also established the National Statistician's Data Ethics Advisory Committee (NSDEC) to ensure transparency around the access, use and sharing of public data for research and statistical purposes, so that it is ethical and for the public good. NSDEC considers projects and policy proposals, which make use of innovative and novel data, from the Office for National Statistics (ONS) and the Government Statistical Service (GSS) and advises the National Statistician on their ethical appropriateness. The Committee has developed a framework of ethical principles and transparent processes to assess proposed research against the principles as illustrated in the table below.⁶⁴

Public good	The use of data has clear benefits for users and serves the public good
Confidentiality, data security	The data subject's identity is protected, information is kept confidential and secure, and the issue of consent is considered appropriately
Methods & quality	The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality
Legal compliance	Data used and methods employed are consistent with legal requirements such as data protection legislation, the Human Rights Act 1998, the Statistics and Registration Service Act 2007 and the common law duty of confidence
Public views & engagement	The views of the public are considered in light of the data used and the perceived benefits of the research
Transparency	The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public

⁶² Article 89 UK GDPR and sections 19 and Part 6 of Schedule 2 to the DPA 2018

⁶³ <https://uksa.statisticsauthority.gov.uk/code-of-practice/> accessed 3 May 2021.

⁶⁴ <https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/national-statisticians-data-ethics-advisory-committee/> accessed 3 May 2021.

Stakeholder experience

As discussed in Case Study 1, stakeholders tended to be of the view that existing legislation broadly caters for the required data processing, including for statistical research. There was, however, some confusion around public release of data in the interests of greater transparency: ‘in terms of release of data we encountered wildly varying opinions as to what counts as small number, identifiable data... we were stopped from releasing datasets that had fewer than 10 cases... [whereas] local authorities had it down to one case. It’s wildly inconsistent’.⁶⁵ Stakeholders commented that guidance with concrete examples would be helpful in this regard.

It was also highlighted that understanding of data ethics could be improved in the mathematical community. One stakeholder was concerned that ‘maths departments... are in deep denial about whether there are any ethical issues at all’.⁶⁶ Arguably the pandemic has emphasised the need for ‘scientific integrity in emergency situations’,⁶⁷ to ensure proper levels of accountability. ‘Having an ethical basis for what we are doing has become really important’, though there was a concern that ‘organisations basically chucked ethics out the window when things got started’.⁶⁸ It was, however, pointed out that data ethics principles must be concrete, so that they can be applied in practice and to avoid ‘ethics theatre’.⁶⁹ This gap could also be addressed as part of the professionalisation of data science by the RSS, which could capture some of the lessons learned during the pandemic: ‘normally people are not very aware of the ethics side and unable to relate but here we have got case studies galore [provided by the pandemic] ... and that is quite helpful’.⁷⁰

‘Doing things ethically is an important enabler. People who put their data into the system need to know that there are processes in place. Ministers need to be assured that data is being used in safe and appropriate ways. It is not an either/or. There is a need to make better use of data but this needs to be done safely and ethically. It is working with researchers to enable them to apply ethical principles to their work... there is a danger that some ethical principles look really good on paper, but are they being used?’

Stakeholder interview D3

LESSONS LEARNED

- Explanatory notices should accompany policy decisions, which address the respective roles of modelling and value judgements in policy decision-making to increase transparency and aid public understanding. This should include understandable information around uncertainty and the use of assumptions, making clear that statistical outcomes do not equate to ‘truth’.
- The gap in understanding between modellers and decision-makers should be addressed as a priority, through interdisciplinary collaboration at leadership level within the public sector and the professionalisation of data science. Data ethics principles, linked to legal requirements and independent oversight, such as those promoted by the UK Statistics Authority, should form part of the professional framework for data science, which could draw on case studies from the pandemic to avoid ‘ethics theatre’.

⁶⁵ Interview with D9

⁶⁶ Interview with D4

⁶⁷ Interview with D4

⁶⁸ Interview with D4

⁶⁹ <https://www.theguardian.com/commentisfree/2021/mar/13/google-questions-about-artificial-intelligence-ethics-doesnt-want-answers-gebru-mitchell-parrots-language> accessed 11 May 2021.

⁷⁰ Interview with D4

CASE STUDY 3

DATA-DRIVEN SUPPORT FOR VULNERABLE RESIDENTS

HEADLINE FINDINGS

- Data has been invaluable in enabling local authorities to identify and support vulnerable residents ‘from a distance’ during the pandemic. This task was made more difficult, however, by data quality issues, leading to concerns that some vulnerable residents would be missed.
- Public bodies have identified lawful reasons for data sharing connected with the emergency situation, with data protection no longer being perceived as a ‘barrier’ to legitimate sharing and use of data. This has enabled a more holistic view of local communities, with stakeholders observing improvements to public services as a result.
- A new ‘conversation’, including an element of direct public engagement, is required to determine the parameters of public acceptability with regard to data use post-pandemic.

CASE STUDY OVERVIEW

During the pandemic, data has proven to be a powerful asset in enabling local authorities to continue providing services to the community in a range of ways (as explored in detail by the Centre for Data Ethics and Innovation⁷¹). One such way is through the sharing, re-purposing, and linking of new and existing datasets across various public sector departments to identify individuals most vulnerable to Covid-19, both clinically and economically, and provide the necessary support.

More specifically, local authorities have deployed health data from central public sector sources (e.g., the NHS shielding patients list and GP data) together with data from local public sector agencies (e.g., school attendance and free school meals data) and existing datasets from across multiple council



departments to identify residents who may be most vulnerable to the effects of the pandemic.

As a result, councils have been able to better direct support to those in need by, for instance, contacting individuals by telephone, email or in person, sending food parcels and organising pharmacy deliveries.

Figure 2 –
Simplified illustration of case study

⁷¹ Centre for Data Ethics and Innovation, ‘Local government use of data during the pandemic’ (2021) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/968515/Local_government_use_of_data_during_the_pandemic.pdf accessed 3 May 2021.

AIMS, OBJECTIVES & BENEFITS

The objective of this data-driven response was to ensure that those most vulnerable would receive the necessary support: ‘the policy behind [it] is: leave no person on their own who might need support, find every person... it was very much about operationalising, targeting where support might be needed’.⁷² This identification extended beyond clinical vulnerability, to include those who were socio-economically vulnerable to the pandemic, and some councils were able to form networks with external support agencies, including food banks, to share data and collectively meet this need.

‘[They] brought data together on who they knew was receiving adult social care support, their homelessness list... the people who needed help moving their bins out because that means they had mobility difficulties which may mean that they also had underlying health conditions... they were triangulating it to try and spot who the potentially vulnerable people were’.

Stakeholder interview G12

Stakeholders consistently shared that the pandemic has been instrumental in demonstrating to local authority leaders the importance of data. ‘It has ‘brought to the fore the power and potential of data in everything local government does’.⁷³ By linking certain existing datasets across departments, councils have been able to work in a way which is more ‘connected up’ and as a result gain a ‘fuller picture’ of members of the community.⁷⁴ ‘It’s been a really successful way of applying data to be more efficient to the general public, and to have a better feeling of understanding of the people you are serving’.⁷⁵ Many stakeholders felt that being able to match data has vastly improved the quality of the service

‘in the past we’ve put ourselves in silos... you’ve got local government and all these health bodies and we’re in silos so we can’t share. Whereas actually if you called and treated us as one department, suddenly we could share data. Nothing has changed, it’s the same people doing the same things with the same systems and we have been making these false walls between data’.

Stakeholder interview G6

they were able to provide to residents; ‘provid[ing] a more holistic view and service instead of different people asking a family the same questions over and over’.⁷⁶ As discussed in the previous case study,⁷⁷ there was a sense that existing ‘barriers’ to data sharing across the public sector, most commonly perceived with regard to information governance, had been overcome, enabling these data-driven benefits to be realised, and a number of stakeholders commented that they are ‘hoping that we can continue that momentum... and not just go back to the old way of doing things. We need to push on really’.⁷⁸

RISKS & CHALLENGES

Data quality: Naturally, one of the central concerns of stakeholders was a worry that they would miss people who needed help, in part due to accuracy issues in the data. As explored in Case Study 1, stakeholders reported variable quality in the data received, with data being received ‘in chaotic forms, [with] no standard way that it’s reported, no reassurance that a vulnerable person has the right

⁷² Interview with G12/13

⁷³ Interview with G12/13

⁷⁴ Interview with G15/16

⁷⁵ Interview with G12/13

⁷⁶ Interview with G7/8

⁷⁷ And further reflected on below in the ‘Law & Governance’ section

⁷⁸ Interview with G6

address'.⁷⁹ Address data, in particular, received from GPs was often incomplete: 'because this person may be in their house and... maybe had no recourse to getting food... [we] were literally sending out people with nothing more than a post code, trying to ask neighbours if they knew this person or could track them down... One high rise block of flats has one post code so it could be hundreds of dwellings' (interview G12/13).⁸⁰ Local authorities appeared to have varying degrees of success in linking their data in view of the quality issues, though a number of stakeholders highlighted the use of unique property reference numbers (UPRNs) as being of great importance in this regard.

Operational capabilities: As one stakeholder observed, 'everything in local government varies massively' (interview G12/13). Accordingly, stakeholders spoke of a range of different approaches to working with the data. Some were 'more sophisticated', incorporating data into their customer relationship management (CRM) systems while others relied heavily on the use Microsoft Excel spreadsheets for working with, and sometimes sharing data, though it was acknowledged that ideally, 'this is not the format you should be using'.⁸¹ Predominantly, however, stakeholders reported that the process for obtaining data from central government was onerous. 'There was never an automatic feed of data made available. No matter how sophisticated a local authority was, there was no API.⁸² So basically, someone had to be there all the time to download – every day and at weekends - ...to keep it going'.⁸³

Skills: Some stakeholders also highlighted a skills gap in local government around data literacy which may hinder progress in the future use of data: 'our Strategic Director ... [is] not very confident when he talks on the subject and that comes through as someone in that domain, [though] he understands the need to have that drive'.⁸⁴

Privacy: Again, privacy was highlighted as a key concern, together with the risk of 'mission creep' regarding the use or sharing of data beyond its original purpose. In a local government context, however, a number of stakeholders held the view – which was often supported by qualitative research within the community - that there was a pre-existing assumption amongst the public that their data would be used in this way: 'the most recent [consultation] was with young people and they were surprised that before Covid-19 we weren't doing these sorts of data sharing activities. They assumed it was already going on'.⁸⁵ As will be explored below, however, stakeholders felt it important to take steps to engage with the public regarding future uses of data so as 'not to make assumptions about that public licence and... to make sure it's maintained'.⁸⁶

⁷⁹ Interview with G12/13

⁸⁰ Interview with G12/13

⁸¹ Interview with G11

⁸² API is the acronym for Application Programming Interface. It is a software intermediary which allows two applications to communicate.

⁸³ Interview with G12/13

⁸⁴ Interview with G11

⁸⁵ Interview with G14

⁸⁶ Interview with G17

LAW, REGULATION & GOVERNANCE

Legal frameworks

As data is being used by local authorities to identify individuals who require additional support, the Data Protection Act 2018 and UK GDPR will apply. Again, therefore, certain principles and justifications for use of that information apply, with particularly stringent justifications required where the data includes information about an individual's ethnic or racial origin or their health, for example.⁸⁷ A common law duty of confidence may apply to confidential information disclosed to health and social care professionals, though again the 'COPI notices'⁸⁸ will apply until 30 September 2021 to enable the sharing of this data without obtaining consent. Public authorities (including local authorities) must also respect individuals' rights to respect for private and family life pursuant to the European Convention on Human Rights (ECHR).⁸⁹ The ECHR thus requires public authorities such as local authorities and health bodies to consider carefully before sharing individuals' information: whether they can do so in accordance with the law; why such sharing is necessary; and what information needs to be shared to achieve such aim, with no more information being shared than necessary. In justifying disclosure, a local authority might, for example, rely upon their obligation to ensure that the right to life⁹⁰ is protected by law.

The requirements imposed by the UK GDPR and ECHR to act in accordance with the law means that information must be shared in accordance with all relevant legal regimes including the DPA and any obligations under the duty of confidentiality. Where the body sharing information is a public body there must also be legal authority for the sharing authority to share such information; most public authorities (i.e., local authorities, police, health bodies) derive their powers from legislation and must act within those powers. Before sharing any information, such authorities should, therefore, identify the relevant legislation to determine whether they have an express obligation or an express power to do so. Where this does not exist, power to share information with another public authority may be implied in legislation. This may be the case, for example, where information sharing is incidental to, or necessary, to enable the organisation to undertake activities permitted by statute. In the health and social care context, data sharing is provided for by the Health and Social Care Act 2012, which imposes an obligation upon local authorities to take appropriate steps for improving health of people in its area. The Care Act 2014 similarly imposes obligations upon local authorities to promote individual wellbeing, to prevent or delay the need for institutionalised care or support, and to exercise

Human rights:

Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) affords a right to respect for private life. Art. 8 covers information which individuals can legitimately expect not to be gathered, published, stored or used without their consent, including medical data (*Z v Finland* (1988) 25 EHRR 371); and personal data more generally (*Satakunnan Markkinapörssi Oy and Satamedia Oy v Finland App no 931/13*). Art. 8 is a qualified right with which the state may interfere in certain circumstances where necessary in a democratic society and in accordance with the law. This includes in the interests of public safety, the protection of health or morals, and for the protection of the rights and freedoms of others. The requirement that interference is necessary entails a test of proportionality; interference must be no more than necessary to meet the stated aim.

⁸⁷ UK GDPR Articles 5, 6, 9.

⁸⁸ NHS Act 2006, s.251; and the Health Service (Control of Patient Information) Regulations 2002 (COPI)

⁸⁹ Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR) art. 8; Human Rights Act 1998, s 6

⁹⁰ ECHR art. 2

its functions ‘with a view to ensuring the integration of care and support provision with health provision and health-related provision.’ In relation to this case study, it may therefore be that the type of data sharing and processing undertaken would be lawful for pandemic-related purposes, though this may not extend to other purposes beyond the pandemic.

Equality:

The Equality Act 2010 sets out the public sector equality duty (s.149) which requires, amongst other things, that public authorities have due regard to the need to advance equality of opportunity between people who share a protected characteristic and those who do not. The protected characteristics (s.4) are: age; disability; gender reassignment; pregnancy or maternity; being married or in a civil partnership; race; religion or belief; sex; or sexual orientation. In particular, in relation to people with protected characteristics, public authorities must take steps to remove or minimise disadvantage; meet the specific needs where they differ from the needs of other people; and encourage participation in public life/activities where participation is disproportionately low. The Act also prohibits direct and indirect discrimination by providers of services to the public based on a protected characteristic.

Public sector organisations also have several obligations under equality legislation, meaning that considerations of equality should be incorporated into the day-to-day business of public authorities. The public sector equality duty is a legal duty and should not be regarded as a ‘tick-box’ exercise, but rather a fundamental part of the decision-making and policy-making processes. Therefore, it is important for public authorities to engage with those affected by their decisions in order to understand the potential impact.

In addition, public sector services must not directly or indirectly discriminate based on a protected characteristic. Stakeholders should therefore be particularly mindful of the risk of discrimination presented by demographic profiling activities, and the potential correlation with protected characteristics, such as ethnicity, for example. This obligation also includes an anticipatory duty on service providers to make reasonable adjustments to avoid disabled people being placed at a ‘substantial disadvantage’ compared to non-disabled people when using or accessing the services.

Both direct and indirect discrimination is covered by the Act – indirect discrimination applies to a provision, criteria or practice that might appear neutral but in fact disadvantages persons with a protected characteristic, and the provision, criteria or practice is not a proportionate means to achieving a legitimate aim. In this regard, one stakeholder did note that the way in which the public is informed of data processing activities - primarily through the publication of a privacy notice online – is potentially problematic and questioned whether this is ‘good enough or fair enough’: ‘do our most vulnerable citizens have access to the website?’⁹¹ This small example may arguably disadvantage certain groups, such as the elderly or those who are visually impaired, and therefore require justification.

I think one thing that was really interesting, everything was set up really quickly and it showed just how fast we could do something. Normally we spend months, sometimes it takes years to agree ISAs [information sharing agreements] and send them back and then it takes another 10 years for everyone to be ready to sign it, by which point is ready to review again. Now I can put a Covid ISA in place in a day.... So, it showed that IG (information governance) wasn't a barrier and data protection isn't a barrier: as long as the legislation is there to do it, we can do it, and if you've got good effective knowledgeable data protection officers, they could do it really quickly. ...again, they have done data protection impact assessments really quickly. So that was something that I think we've learned, how quickly we can do things.

Stakeholder interview G6

⁹¹ Interview with G18/19

Stakeholder experience

As discussed in Case Study 1, stakeholders felt that the pandemic had provided the impetus to overcome the perceived barriers to data sharing, particularly in the public sector. Many stakeholders believe the overly 'risk-averse' interpretation of data protection legislation has been replaced by a more facilitative, 'human-centred' attitude to information governance.⁹² This has enabled local authorities to act in a quicker, more efficient manner in the interests of responding to the pandemic: 'recently where someone came back from [abroad] into their local area - who Public Health England needed to be contacted - we have been able to message council tax colleagues and obtain the information from them, so that we can contact the person... Previously we always just got [a response of] "we can't possibly give you that information"'.⁹³ Stakeholders were eager to try to 'build on' these successes and ensure that this momentum is maintained for future ways of working, particularly as more senior local authority leaders are beginning to see 'the benefits of this resource they have in their organisation and [are] understanding it in ways perhaps they hadn't before'.⁹⁴ Many stakeholders were however mindful of the fact that there will be a number of obstacles, particularly budget constraints and competing priorities: 'It's something that always drops to bottom of the pack though when you're looking at protecting children (for example), there are more important things to worry about than having a team of analysts and business-people in the background'.⁹⁵

I'm careful about making assumptions, because we do have a particular licence to continue to do things safely, so it has been important that we don't omit that step because it's vital that we maintain public confidence in how we are using data both during the pandemic and afterwards. but I also think that during the pandemic we have both legally and in terms of public opinion I think we have a certain licence to move at a different pace. I think when looking to the future we're being very careful not to make assumptions about that public licence and need to make sure it's maintained; we are planning to do some further public engagement, looking carefully at what frameworks we can create that enable use of public data for innovation and creating public benefit.

Stakeholder interview G17

Fundamentally, it was also questioned whether the momentum *should* be maintained in a non-emergency, 'business as usual' situation. 'There will be some that will say [we need to] keep the momentum but actually, we should respect the right to return back to our previous position and start a new conversation, which I think is the proper way to do it'.⁹⁶ In this regard, several stakeholders felt that further public engagement was required to ensure that assumptions weren't being made with regard to the public acceptability, to promote transparency and to maintain public confidence.

Addressing this issue, one local authority stakeholder had commenced a piece of work prior to the Covid-19 pandemic involving community groups to create a formal 'data charter'. This will govern internal data ethics processes alongside existing information governance.⁹⁷ More broadly, stakeholders did incorporate some form of ethical review into their information governance processes, with varying degrees of formality. For some stakeholders this was a formal step to be completed as

⁹² Interview with P2

⁹³ Interview with G15/16

⁹⁴ Interview with G12/13

⁹⁵ Interview with G12/13

⁹⁶ Interview with G11

⁹⁷ Interview with G7/8

part of the Data Protection Impact Assessment, requiring consideration of specific key issues such as: the effect on residents (including benefits and harms), the possibility of data bias and its impact, limitations in the data, proportionate use, and transparency. For other stakeholders, this was more of an informal consideration: ‘we often throw in the question... does it feel right? If it doesn't feel right, what are we missing here? And then quite often you might think about the Human Rights Act aspect, that maybe you haven't gone into enough detail on. Or a gut feeling that something doesn't feel right. If they were doing that with my data, I wouldn't be happy - why not?’.⁹⁸ There was some concern, however, that ‘data ethics as an overall concept isn't widely understood [at the council] at all, and specifically just because we have access to this data and can link all this data, is it appropriate for me to do it? It's a conversation’.⁹⁹

LESSONS LEARNED

- Investment and resource should be directed to address shortcomings in public sector data as a priority. Going forward, a **data standards framework** should be agreed and implemented to address inconsistencies and enable interoperability of public sector data.
- An examination of the variability in **operational capabilities** across local authorities regarding data use would be beneficial. This should include evaluation of any data literacy **skills gaps** to be addressed through additional training or professional resource (for example using the RSS professional standards framework for data science) and the formalisation of ethical review. Any form of ethics review should, however, be based on a meaningful framework of concrete principles which can be operationalised to avoid ‘ethics theatre’. This should incorporate an additional step for independent advice or oversight.
- A new public ‘conversation’ is needed regarding the post-pandemic use of data, involving, at the very least, **public engagement**, but preferably, public consultation to determine the parameters of public acceptability.

⁹⁸ Interview with G6

⁹⁹ Interview with G11