

A woman with dark hair tied back, wearing a light blue surgical mask and a dark jacket, is shown in profile, looking towards the left. The background is a blurred city street with cars and buildings.

The
Alan Turing
Institute

Data science and AI in the age of COVID-19

Reflections on the response of the UK's
data science and AI community to the
COVID-19 pandemic

Executive summary

This report summarises the findings of a series of workshops carried out by The Alan Turing Institute, the UK's national institute for data science and artificial intelligence (AI), in late 2020 following the 'AI and data science in the age of COVID-19' conference. The aim of the workshops was to capture the successes and challenges experienced by the UK's data science and AI community during the COVID-19 pandemic, and the community's suggestions for how these challenges might be tackled.

Four key themes emerged from the workshops.

First, the community made **many contributions** to the UK's response to the pandemic, via national organisations, research institutes and the healthcare sector. Researchers responded to the crisis with ingenuity and determination, and the result was a range of new projects and collaborations that informed the pandemic response and opened up new areas for future study.

Second, the single most consistent message across the workshops was the importance – and at times lack – of robust and timely data. Problems around **data availability, access and standardisation** spanned the entire spectrum of data

science activity during the pandemic. The message was clear: better data would enable a better response.

Third, issues of **inequality and exclusion** related to data science and AI arose during the pandemic. These included concerns about inadequate representation of minority groups in data, and low engagement with these groups, which could bias research and policy decisions. Workshop participants also raised concerns about inequalities in the ease with which researchers could access data, and about a lack of diversity within the research community (and the potential impacts of this).

Fourth, **communication** difficulties surfaced. While there have been excellent examples of science communication throughout the pandemic, participants highlighted the challenges of communicating research findings and uncertainties to policy makers and the public in a timely, accurate and clear manner.

In this report, we outline the workshop participants' reflections and suggestions relating to each of these themes, with the aim of enabling the data science and AI community to respond better to the ongoing pandemic, and future emergencies.

Contents

Foreword	4
Preface	5
1. Spotlighting contributions from the data science and AI community	7
+ The Turing's response to COVID-19	10
2. Data access and standardisation	12
3. Inequality and exclusion	14
4. Communication	17
5. Conclusions	20
Appendix A. Workshop participants and Organising Committee	22
Appendix B. Workshop themes and reports	24

Edited by



Inken von Borzyskowski
Assistant Professor of Political Science, University College London



Anjali Mazumder
AI and Justice and Human Rights Theme Lead, The Alan Turing Institute



Bilal Mateen
Clinical Data Science Fellow, The Alan Turing Institute; Clinical Technology Lead, Wellcome Trust



Michael Wooldridge
Turing Fellow and Programme Co-Director for Artificial Intelligence, The Alan Turing Institute

Workshop theme leads

Mark Briers (The Alan Turing Institute), Tao Cheng (UCL), Spiros Denaxas (UCL), John Dennis (University of Exeter), Sabina Leonelli (University of Exeter), David Leslie (The Alan Turing Institute), Marion Mafham (University of Oxford), Ed Manley (University of Leeds), Karyn Morrissey (University of Exeter), Deepak Parashar (University of Warwick), Jim Weatherall (AstraZeneca)

Foreword

The COVID-19 pandemic has seen scientific research move into public discourse in unparalleled ways. Across the scientific spectrum, researchers have stepped up, the data science and AI community included, to work alongside clinicians, policy makers and the government at the heart of the response, directly impacting on our daily lives.

Data science and AI is an inherently interdisciplinary community, and our activities at The Alan Turing Institute in response to the pandemic reflect this. Our researchers have developed algorithms to monitor pedestrian density and ensure social distancing on the streets of London; combined NHS datasets to help answer clinical questions about the effects of COVID-19; explored what makes people vulnerable to health-related misinformation; and improved the accuracy of the NHS COVID-19 app. (See page 10 for more details of our response.)

This report has been edited by four researchers with backgrounds spanning AI, data science, public policy, human rights and medicine. They have synthesised the views of 96 attendees to a series of workshops held at the end of 2020, which aimed to provide a snapshot of the uses of data science and AI during the pandemic, and what we as a community can learn from the experience.

While this represents a small part of what will surely be a larger reflection exercise to come, the central findings – issues of data access and quality; inequality among both the research community and wider society; and communication difficulties between experts

and non-experts – contain valuable reflections and suggestions for how the data science and AI community might prepare for future emergencies. Indeed, the Turing has already begun a large-scale project¹ which aims to boost societal, governmental and economic resilience to shocks such as this pandemic.

Ahead of the G7 summit in the UK in June 2021, the leading scientific bodies of the G7 nations (the ‘Science 7’) recently published a call for more ‘data readiness’ in preparation for future health emergencies.² This is a timely amplification of the message in the Turing’s report about the need for increased data access and sharing, at a time when the pandemic continues to have catastrophic impacts around the world.

My thanks to the editors for initiating and delivering this report, and to all the workshop theme leads and participants. We hope it will be a valued contribution to the ongoing discussions about the data science and AI response, alongside notable reports from the Centre for Data Ethics and Innovation,³ the Ada Lovelace Institute,⁴ the Royal Society’s DELVE initiative⁵ and the Royal Statistical Society,⁶ among others.

The Turing looks forward to continuing its role to convene and deliver activities and reflections in response to this pandemic, and in preparation for other crises.

Adrian Smith

Institute Director and Chief Executive
The Alan Turing Institute

¹ <https://www.turing.ac.uk/research/research-projects/shocks-and-resilience>

² <https://royalsociety.org/-/media/about-us/international/g-science-statements/G7-data-for-international-health-emergencies-31-03-2021.pdf>

³ <https://www.gov.uk/government/publications/covid-19-repository-and-public-attitudes-retrospective>

⁴ <https://www.adalovelaceinstitute.org/summary/learning-data-lessons>

⁵ <https://rs-delve.github.io/reports/2020/11/24/data-readiness-lessons-from-an-emergency.html>

⁶ <https://rss.org.uk/statistics-data-and-covid>

Preface

At the end of 2019, a new highly infectious virus, now known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified as the underlying pathogen for a series of unexplained pneumonias (subsequently termed coronavirus disease 2019, or COVID-19), clustered in Wuhan, China. By 30 January 2020, COVID-19 had become so prevalent that the World Health Organisation (WHO) declared it a Public Health Emergency of International Concern – the WHO’s highest level of alarm. As we finish this report in spring 2021, the disease itself has claimed over three million lives globally, with more than 170 million confirmed cases, and many more affected by the impacts of lockdowns and the unprecedented disruption to the global economy. The UK has now been through two waves of the virus, with infections, hospitalisations and fatalities in the second wave exceeding those in the first.

While pandemics appear to have occurred throughout human history, the COVID-19 pandemic is unique in one important respect. It is the first pandemic to occur in the age of data science and AI: the first pandemic in a world of deep learning, ubiquitous computing, smartphones, wearable technology and social media. It is thus unsurprising that governments across the globe, including the UK’s, looked to data to inform their responses and help navigate challenges. The goal was to limit the spread of the disease and its medical, social and economic consequences. As such, the UK government stated that its policies were “guided by the science”, and later that ending lockdowns depended on “data, not dates”.

In response, many members of the UK’s data science and AI community stepped forward, spearheading initiatives that they hoped would assist the domestic and international response. These initiatives came from individual academics, university research groups, the healthcare sector, national institutes and others. They involved not just experts on virology and epidemiological modelling, but also researchers studying, for example, the social and economic consequences of non-

pharmaceutical interventions. The response was remarkable for its breadth of engagement across disciplines, as demonstrated by the range of backgrounds of our workshop participants and the diverse set of insights that they generated.

Goals, origins and structure of the report

This report was commissioned by The Alan Turing Institute with the aim of reflecting on the UK’s data science and AI response to the pandemic. The Turing is the UK’s national institute for data science and AI, and partners with many of the UK’s leading universities and research centres to advance the country’s capacity and competitiveness in these areas, with the overall mission of changing the world for the better.

The Turing’s goals in undertaking this work were twofold:

1. To capture the initiatives and resources that have been developed by the data science and AI community during the pandemic.
2. To gather the experiences and insights of this community during the pandemic – what worked well, what didn’t, and how we as a community could respond better to this pandemic and future emergencies.

The work has its origins in a one-day conference ‘AI and data science in the age of COVID-19,’⁷ which was held virtually by the Turing on 24 November 2020 and featured talks from some of the leading voices in the UK’s response to COVID-19. The free event attracted over 1,700 registrants from 35 countries, from academia, industry, the public sector and the general public.

A series of themed, virtual workshops followed in November and December 2020. The invitation to participate in these workshops was widely circulated within the UK academic community, via social media and the Turing’s network of partners and affiliates. The Turing used a lightweight reviewing process to select the participants, who are listed in Appendix A.

⁷ <https://www.turing.ac.uk/events/ai-and-data-science-age-covid-19>

We took the view that, if the report was to have any credibility, it would be essential to include the views of as diverse a community as possible. The scope of 'data science and AI' for the purposes of this report is therefore deliberately broad, with participants including ethicists, clinicians, mathematicians, policy advisors, and many more.

The eight workshop themes are listed in Appendix B. These themes are themselves a reflection of input from over 20 experts, who formed the Organising Committee for the conference and workshops. All workshops were led by the Centre for Facilitation⁸ together with one or two workshop-specific theme leads who were selected by the Organising Committee to supplement the facilitators' expertise with domain knowledge. For each workshop, the facilitators and theme leads summarised the discussions in a report, which have been lightly edited and linked from Appendix B. These workshop reports provided the 'raw data' for this report. Given the range and complexity of the workshops, it has not been possible to include all of the participants' views and suggestions, but we as editors have tried to draw out the main themes.

Although our remit includes both data science and AI – reflecting the Turing's role as the national institute for these two subjects – the majority of the feedback we received from workshop participants related specifically to data science, and our report reflects this. This chimes with the 'COVID-19 repository and public attitudes' report⁹ published by the Centre for Data Ethics and Innovation (CDEI) in March 2021, which found that "conventional data analysis has been at the heart of the COVID-19 response, not AI". There were certainly innovative applications of AI during the pandemic, but the evidence suggests that more traditional methods of data collection

and analysis were far more widespread. One possible reason for this, suggested by the CDEI report, is that COVID-19 was such a new phenomenon that there was a lack of data with which to train AI algorithms.

We also note that while we as editors have done our best to report the suggestions of workshop participants as accurately as possible, in some cases we have applied light editing. This especially applies to the suggestions around data access in Chapter 2, where we have taken care to put the participants' aspirations for more accessible and open data in the context of the complex legal and ethical obligations surrounding access to sensitive national health, economic and social information. Data access is clearly an important issue which needs to be addressed, and there are many steps to enabling responsible data flows.

One final and important word. Our report is preliminary in the sense that we hope it will serve as a starting point for a more comprehensive and systematic review of the uses of data science and AI during the pandemic. The report was originally conceived in summer 2020, when daily reported COVID-19 fatalities in the UK were down to single figures, and there was some hope that the virus might be largely contained in a single wave. Work on this report continued during the November 2020 to spring 2021 lockdowns when the disease was again prevalent. In this sense, many (but not all) of the observations made here relate to the first wave of the pandemic in the UK, from January to summer 2020, prior to the onset of the UK's second wave. We hope that the report will be of value to healthcare professionals, data science and AI researchers, and policy/decision makers as we continue to manage the ongoing pandemic.

1. Spotlighting contributions from the data science and AI community

A key goal of the workshops was to document the contributions from the UK's data science and AI community in responding to the pandemic. In this chapter, we summarise the contributions highlighted by the workshop participants, which demonstrate some positive aspects of the data science and AI response: increased data sharing, collaboration across disciplines, and the development of new datasets, repositories and initiatives that mobilised COVID-19 research.

We emphasise that these spotlights are nothing more than that: spotlights. The workshops provided a snapshot of the community's response, but there were many other contributions that were not covered in the workshops. An overview of The Alan Turing Institute's key COVID-19 projects can be found on page 10 of this report, and further examples of the community's response can be found in resources such as the CDEI's COVID-19 repository.¹⁰ Finally, we acknowledge that there may be some bias in this chapter, with participants more likely to mention projects that they are affiliated with.

Noteworthy datasets and databases

The UK government's action to issue statutory regulators such as NHS Digital with a Control of Patient Information (COPI) notice¹¹ to make confidential patient data available to researchers and policy makers for COVID-19-specific purposes had far-reaching consequences, and participants noted the positive effect this had on data sharing.

For example, it paved the way for a new statistical analytics platform called **OpenSAFELY**,¹² a collaboration between the University of Oxford's DataLab, the London School of Hygiene & Tropical Medicine, and

several healthcare technology companies. OpenSAFELY provides secure access to a database of over 58 million NHS patient records, allowing researchers to answer urgent clinical and public health questions related to COVID-19. Moreover, for the first time it separates the development of the analysis code from the actual data (which never leave the servers of the electronic health record service provider or NHS Digital), and in doing so guarantees fully open and reproducible code. This is a noteworthy development that occurred largely due to the permissive nature of the regulatory environment necessitated by the pandemic.

Another new initiative was the **COVID-19 Genomics UK Consortium (COG-UK)**,¹³ which has sequenced more than 490,000 SARS-CoV-2 virus genomes to date, providing important information on viral transmission and the emergence of new variants. COG-UK was achieved through the coordination of several organisations (NHS, the UK's public health agencies, the Wellcome Sanger Institute and various academic institutions). Having those sequences available from the earliest stages of the pandemic was key to pushing forward the development of vaccines.

Existing systems and databases from **ISARIC** (International Severe Acute Respiratory and emerging Infection Consortium)¹⁴ and **ICNARC** (Intensive Care National Audit & Research Centre)¹⁵ were repurposed and expanded for the pandemic, which kept costs down and enabled the data science community to act more quickly. Similarly, linked data from GPs and NHS trusts were available to researchers through the pre-existing **Discovery East London**¹⁶ platform. Participants also highlighted the utility of aggregate data on critical care patients collected by **CHES**

⁸ <https://www.centreforfacilitation.co.uk>

⁹ <https://www.gov.uk/government/publications/covid-19-repository-and-public-attitudes-retrospective>

¹⁰ <https://www.gov.uk/government/publications/covid-19-repository-and-public-attitudes-retrospective>

¹¹ <https://digital.nhs.uk/coronavirus/coronavirus-covid-19-response-information-governance-hub/control-of-patient-information-copi-notice>

¹² <https://opensafely.org>

¹³ <https://www.cogconsortium.uk>

¹⁴ <https://isaric.org>

¹⁵ <https://www.icnarc.org>

¹⁶ <https://www.eastlondonhcp.nhs.uk/ourplans/discovery-east-london-case-study.htm>

(COVID-19 Hospitalisation in England Surveillance System),¹⁷ which was adapted from the UK Severe Influenza Surveillance System by Public Health England.

Participants noted that a major breakthrough in the data science community's response to the pandemic was the opening up (and sharing) of various **mobility datasets**¹⁸ by commercial providers, such as **Cuebiq**,¹⁹ **Google**,²⁰ and **Facebook's Data for Good**.²¹ These enabled the detailed spatial analysis and large-scale simulation of viral transmission that was important for generating evidence for the government and the public on the potential and actual impacts of lockdowns. Public transport data from **Transport for London** and the **Department for Transport** were cited as central to arguments about the effectiveness of the first lockdown. The ability to monitor data in real time through **dashboards** (e.g. **i-sense COVID RED**,²² **Evergreen**,²³ **GOV.UK**²⁴ and **Public Health Scotland**²⁵) also helped the data science community to respond and report trends to the public.

Workshop participants praised the work of the **Office for National Statistics (ONS)** in supplying data during the pandemic, including its work on deaths stratified by ethnicity,²⁶ the social impacts of COVID-19,²⁷ and how people spent their time during lockdown.²⁸ There was also praise for the work of **Health Data Research UK (HDR UK)**, specifically its Innovation Gateway,²⁹ which provides information about the datasets and tools held by members of the UK Health Data Research Alliance,³⁰ facilitating the navigation of multiple resources relevant to COVID-19.

Participants also cited the **DECOVID**³¹ project, which aimed to create a detailed, frequently updated database of anonymised patient health data during the pandemic. The project was initiated during the early stages of the pandemic, and involved researchers from The Alan Turing Institute and four other founding institutes. The initial funding, diverted from an existing Turing grant from the Engineering and Physical Sciences Research Council (EPSRC), covered the transfer and combination of data from two NHS trusts, plus the data analysis planning. The work continues with in-kind contributions from researchers around the UK, who are analysing the data (now covering 185,000 patients) to shed light on four key questions, including when to put critical COVID-19 patients onto a ventilator, and how patients with long-term health conditions are affected by COVID-19. Ethicists in the Turing's public policy programme have been embedded within the research teams from the start of the project to ensure algorithms are implemented to the highest standards of transparency and bias mitigation. The first results from DECOVID are expected later in 2021.

Finally, participants noted regional data management systems that had proved valuable. The **SAIL Databank**³² data linkage service was instrumental in redeploying existing datasets to support pandemic efforts in Wales. And **DataLoch**³³ is a repository of all routine health and social care data for the Edinburgh and South East Scotland Region, which provides data to a range of researchers to address COVID-19-related questions.

Other initiatives

The **Royal Society** was highly active during the pandemic, using its convening power to support the UK's pandemic response through two initiatives. First, **RAMP (Rapid Assistance in Modelling the Pandemic)**³⁴ brought together individuals with modelling expertise from a diverse range of disciplines to support the COVID-19 pandemic modelling effort. For example, workshop participants highlighted a collaboration between RAMP's Urban Analytics team and Improbable (a UK technology company), which developed a 'micro-simulation' model of the spread of COVID-19 based on highly realistic synthetic data of people's daily activity – i.e. where they go (home, shops, school, work) and for how long. The project's main outcome was a model of Devon's entire 800,000-person population, which allowed researchers to compare the impact of different intervention scenarios at a local level.³⁵

Second, **DELVE (Data Evaluation and Learning for Viral Epidemics)**³⁶ is a multidisciplinary group of researchers convened by the Royal Society to contribute data-driven analysis to the UK's pandemic response, providing input through SAGE, the government's Scientific Advisory Group for Emergencies. Outputs from the group include reports on the efficacy of face masks in tackling COVID-19³⁷ and the risks associated with pupils returning to schools in September 2020.³⁸

Workshop participants also noted positive examples of **public engagement** from the data science and AI community, such as the contributions of Professor Christina Pagel and Professor Devi Sridhar, in helping to improve understanding of the pandemic and policy interventions. Participants also highlighted the work by the **Ada Lovelace Institute** as a good example of how to effectively engage the public during lockdown. In May and June 2020 (during the first UK lockdown) the Ada Lovelace Institute worked with Traverse,³⁹ Involve,⁴⁰ and Bang the Table⁴¹ to conduct a rapid online deliberation with 28 members of the public to explore attitudes to the use of COVID-19-related technologies, such as contact tracing apps, in exiting lockdown.⁴² The aim of the project was to test this new engagement methodology, which could be used when face-to-face public deliberations are not possible, and it resulted in a report on how to build trust in technologies developed in response to the pandemic.⁴³

Lastly, although not mentioned by workshop participants, the editors would like to highlight The Trinity Challenge⁴⁴ as a particularly exciting initiative to come out of the pandemic. This coalition of organisations from business, academia and the social sector is offering an award pool of £10m for ground-breaking, data-driven solutions that will help countries to better identify, respond to, and recover from outbreaks of disease.

“...some positive aspects of the data science and AI response: increased data sharing, collaboration across disciplines, and the development of new datasets, repositories and initiatives that mobilised COVID-19 research.”

17 <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>

18 http://theodi.org/wp-content/uploads/2021/04/Data4COVID19_0329_v3.pdf

19 <https://www.cuebiq.com>

20 <https://www.google.com/covid19/mobility>

21 <https://dataforgood.fb.com/docs/covid19>

22 <https://covid.i-sense.org.uk>

23 <https://www.evergreen-life.co.uk/covid-19-heat-map>

24 <https://coronavirus.data.gov.uk>

25 https://public.tableau.com/profile/phs.covid.19#!/vizhome/COVID-19DailyDashboard_15960160643010/Overview

26 <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronavirus-covid19relateddeathsbyethnicgroupenglandandwales/2march2020to15may2020>

27 <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandthesocialimpactsongreatbritain/4december2020>

28 <https://www.ons.gov.uk/economy/nationalaccounts/satelliteaccounts/bulletins/coronavirusandhowpeoplespent-theirtimeunderrestrictions/28marchto26april2020>

29 <https://www.healthdatagateway.org>

30 <https://ukhealthdata.org>

31 <https://www.decovid.org>

32 <https://saildatabank.com>

33 <https://www.ed.ac.uk/usher/dataloch>

34 <https://royalsociety.org/topics-policy/Health%20and%20wellbeing/ramp>

35 <https://www.improbable.io/blog/improbable-synthetic-environment-technology-accelerates-uk-pandemic-modelling>

36 <https://rs-delve.github.io>

37 <https://royalsociety.org/news/2020/05/delve-group-publishes-evidence-paper-on-use-of-face-masks>

38 <https://royalsociety.org/news/2020/07/delve-opening-schools-should-be-prioritised-report>

39 <http://traverse.org.uk>

40 <https://www.involve.org.uk>

41 <https://www.bangthetable.com>

42 <https://www.adalovelaceinstitute.org/project/rapid-online-deliberation-on-covid-19-technologies>

43 <https://www.adalovelaceinstitute.org/report/confidence-in-crisis-building-public-trust-contact-tracing-app>

44 <https://thetrinitychallenge.org>

The Turing's response to COVID-19

Since the beginning of the pandemic, The Alan Turing Institute has been working to tackle the spread and effects of COVID-19. Here are some of our key projects – visit our [dedicated webpage](#) for more info.

Helping London to navigate lockdown safely

As London locked down in spring 2020, a team in the Turing's data-centric engineering programme began [Project Odysseus](#) to monitor levels of activity on the city streets. Working with the Greater London Authority (GLA) and Transport for London (TfL), the researchers fed their algorithms with data from traffic cameras and sensors to provide anonymised, near-real time estimates of pedestrian densities and distances. TfL used this tool during the pandemic's first wave to make numerous interventions to keep people socially distanced, such as moving bus stops, widening pavements and closing parking bays.

Combining data from NHS trusts

Initiated by the Turing and four other partners, the [DECOVID](#) project has created a detailed database of anonymised patient health data. A major breakthrough has been the transfer and combination of data from two NHS trusts, covering 185,000 patients. The work continues with in-kind contributions from researchers around the UK, who are analysing the data to shed light on four key clinical questions, including when to put critical COVID-19 patients onto a ventilator, and how patients with long-term health conditions are affected by COVID-19. The first results are expected later in 2021.

Modelling the spread of COVID-19 in urban areas

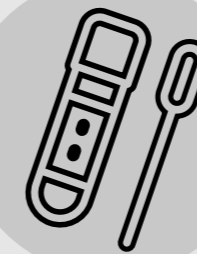
The Turing led the [urban analytics workstream](#) of the Royal Society's Rapid Assistance in Modelling the Pandemic (RAMP) initiative. Using highly realistic data of people's daily activities in towns and cities, the workstream developed a model that simulates COVID-19 transmission at an individual level. A demonstration model of Devon's entire population allowed researchers to compare the impact of different lockdown strategies. The team is now scaling up its model to a national level, and is in dialogue with policy makers about using it to inform decision-making in this pandemic and future health emergencies.

The Alan Turing Institute



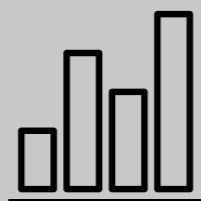
Understanding vulnerability to health-related misinformation

In response to the growing problem of misinformation around COVID-19, the Turing's public policy programme launched a [project](#), funded by The Health Foundation, to understand who is most vulnerable. They [found](#) that people with lower numerical, health and digital literacy tend to fare worse at assessing health-related statements, which suggests that developing people's literacies has the potential to make a big difference to their ability to identify misinformation. The researchers are now hoping to feed into government policy-making around measures to counter the problem.



Estimating positive COVID-19 test counts

The Turing has [partnered](#) with the Royal Statistical Society to provide modelling and machine learning expertise to the UK government's Joint Biosecurity Centre. One of the key outputs so far is a [statistical model](#) that uses incoming COVID-19 test data to estimate ('nowcast') the total number of positive tests in local authorities. It can take up to five days for PCR tests to be processed and reported, so these nowcasts will give authorities an earlier picture of the disease's spread and aid decision-making.



Improving the accuracy of the NHS COVID-19 app

Turing researchers have played an integral role in the [development of the NHS COVID-19 app](#), providing technical advice to the Department of Health and Social Care. The researchers improved the algorithm behind the app's Bluetooth Low Energy contact tracing technology, so that it more accurately calculates the risk that the phone's user has been in contact with a COVID-positive person. A [statistical analysis](#) published in Nature estimated that the app prevented around 600,000 COVID-19 cases in October-December 2020 alone, due to people self-isolating following contact with an infected person.



Building resilience against future crises

A new two-year, multidisciplinary project at the Turing is seeking to better understand how interconnected health, social and economic systems are affected by shocks such as the pandemic. The ['Shocks and resilience'](#) project will develop a coupled epidemiologic and socio-economic model of the spread and societal effects of COVID-19, as well as more generalised models for other complex, socio-economic systems. The overall aim is to produce data, methodologies and tools to enable policy makers to make better informed decisions, boosting the resilience of local and national governments against future shocks.

2. Data access and standardisation

Despite the contributions highlighted in Chapter 1, workshop participants articulated a number of challenges that hindered the efforts of the data science and AI community to contribute to the national effort. The most prominent of these were around data access and standardisation – not having access to the desired data, or having data that were not suitably formatted or documented.

Data access

While some members of the data science and AI community had easy and rapid access to relevant data, several participants commented that many researchers were limited in their contributions because access to data was often restricted, inconsistent or slow.

For example, some hospitalisation data were not available early in the pandemic, and geographically disaggregated data for local analysis and crafting of solutions (e.g. local lockdowns) were also not available to all relevant academic groups. Further, local-level economic data on employment and production by industry sector were not accessible or existent (and possibly often both), making the economic impact of proposed lockdown measures on different industries, such as the hospitality sector, more difficult to assess. More broadly, systematic data on non-pharmaceutical interventions (social distancing, mask wearing, lockdowns), and particularly compliance with such interventions, have not been made appropriately available, making it difficult to measure the impact of these policies on behaviour.

Consequently, insights from the data science and AI community were not as informative and robust as they could have been. It is unclear how much the quantity and quality of research could have been increased if a larger share of the community had been able to contribute its

expertise to the national effort. However, there was a strong perception in the workshops that providing researchers with better data access and generating a more level playing field regardless of affiliation and connections (see also Chapter 3) would likely have helped better address the knowledge and policy challenges during the pandemic.

Data standardisation

Many participants also noted a lack of data standardisation as a significant hurdle for data scientists during the pandemic. Different data standards and codification of metadata, and lack of dataset documentation, meant that data were difficult to find, link and assess in terms of missingness and biases, limiting the scope of and confidence in analyses.

For example, ideally it would have been possible to link datasets from different studies, but many studies and consortia funded by UK Research and Innovation (UKRI, e.g. ISARIC4C⁴⁵ and PHOSP-COVID⁴⁶) are relatively standalone. It would also have been beneficial to standardise codes across datasets, including variables for vulnerable and underserved groups, ethnicity and socio-economic status. Better data standardisation would have enabled data scientists to more easily find, understand, link and triangulate information from different datasets in order to better answer policy-relevant questions and to allow testing for individual-level or context effects on outcomes (e.g. healthcare, employment, public information). As editors, we acknowledge that this aspiration for linked and standardised datasets will be difficult to achieve in the near-term: different research communities have different standards, formal or informal, different modes of working, and often use similar terminology with quite distinct meanings. Nevertheless, the value of such datasets is clear, and much more could clearly be done.

Suggestions

Reflecting on these data challenges, participants made the following suggestions for the data science and AI community:

- Aspire to a **research culture in which data are shared as openly as legal and ethical obligations permit**, with central repositories, or ‘data lakes’, for cleaned and anonymised data (or weblinks to data) ready for analysis. Document which data exist (to avoid duplication of effort) and make them widely available.
- Data repositories should **signpost to open datasets** and, for non-open data, to the details of data sharing agreements and protocols for securely accessing the datasets. This can ensure the wider provision and availability of shared datasets and equitable data access.
- **Investigate ways in which access to sensitive data (e.g. from the NHS) may be enabled** while respecting professional, ethical and legal obligations surrounding the data. This will require developing new ways to allow researchers to securely access personal data, perhaps using differential privacy or federated learning techniques. This in turn presents new research challenges around making these latter

techniques work robustly, and at scale, with real-world data, while also considering data governance principles and practices.

- **Automate data collection systems.** Data collection is resource-intensive, so automated data collection could reduce the reliance on frontline staff to record data using manual systems. This aspiration needs to be set against the obvious need to ensure the rights of individuals and communities to privacy and agency.
- **Encourage more data sharing agreements.** Develop more general data sharing agreements as opposed to those for narrow purposes/groups of users to reduce barriers and delays.
- **Develop protocols for minimum standardisation** of data fields (including some baseline data to always be included) to ease linking of datasets and comparable metadata. As part of this, consider following the ‘FAIR Guiding Principles for scientific data management and stewardship’⁴⁷ Findability, Accessibility, Interoperability and Reusability.
- **Conduct a stress test** of the new standardisation specifications by using regular events (such as the annual influenza season) to prepare for the next pandemic.

“Aspire to a research culture in which data are shared as openly as legal and ethical obligations permit, with central repositories, or ‘data lakes’, for cleaned and anonymised data ready for analysis.”

⁴⁵ <https://isaric4c.net>

⁴⁶ <https://www.phosp.org>

⁴⁷ <https://www.go-fair.org/fair-principles>

3. Inequality and exclusion

The COVID-19 pandemic has brought societal inequality into sharp focus, with the disease having a much greater impact on some groups than others. In England and Wales during the first wave of the pandemic, nearly all minority ethnic groups had higher mortality rates than the White ethnic population.⁴⁸ There were also marked socio-economic differences, with mortality rates in the most deprived areas around double those in the least deprived areas.⁴⁹

Alongside these broader societal issues, problems of inequality and bias relating to data science and AI have also been highlighted by the pandemic. There have been long-term concerns about the potential of these research areas to marginalise certain groups, but the rise during the pandemic of new datasets, data capture/analysis methods and data-driven technologies such as contact tracing apps, plus ongoing discussions around immunity and vaccine passports, has meant that these issues are more pertinent than ever. We need ongoing scrutiny to ensure that any tools and technologies developed by the data science and AI community respond to people's health, socio-economic and security needs in a fair and equitable manner.

Inequality and exclusion were common themes in the workshops, and the participants' comments fell into two main categories:

- Inequality and exclusion within society
- Inequality and exclusion within the research community

Inequality and exclusion within society

Participants felt that the potential of the data science and AI community to help understand the impact of COVID-19 on different ethnic and socio-economic groups was not fully realised due to several challenges.

Many participants commented on a lack of relevant or consistent data. While ONS and other public data providers released up-to-date COVID-19 data on minority groups, many data sources had gaps at both the national and local

level. For instance, participants cited problems around the standardisation of capture of socio-economic status and ethnicity.

While participants acknowledged that a robust and rigorous system exists for obtaining epidemiological data via institutions such as Public Health England, they noted that such a system does not exist for wider economic, mobility and socio-economic data, and that this gap hampered socio-economic research in the early weeks of the pandemic. Participants also noted a lack of sufficiently granular, local-level data on transmission of COVID-19 across communities, which hid some of the social inequalities associated with the initial phase of the pandemic.

"The COVID-19 pandemic has brought societal inequality into sharp focus."

Participants raised concerns that the scramble to use existing datasets, or to rapidly create new ones, ran the risk of proceeding without due regard for sampling biases. These biases, such as insufficient representation of minority groups, could be 'baked into' datasets, and may be due to systemic discrimination, structural inequalities and/or data collection constraints. This could in turn lead to biased research and policies that exacerbate pre-existing inequalities. Data availability, detail and representativeness during clinical trials also had shortcomings, which inhibited the inclusion of patient subgroups, and may have made it more difficult to assess, for example, the efficacy of COVID-19 treatments and vaccines in minority ethnic groups.

Participants also noted significant inequality and exclusion challenges around the use of mobility, mobile device and other digital data in behavioural analysis. For instance, traditionally underrepresented groups also tend to be scarce in digital data – especially those on the 'invisible' side of the digital divide, who might

not have access to the internet or a computer, or who have not engaged with initiatives due to historical distrust within their community.

Finally, participants commented that the data science and AI community had been insufficiently engaged with disadvantaged groups on what COVID-19-related research would be useful for them, with the Ada Lovelace Institute cited as one of the few organisations talking to these groups.

Inequality and exclusion within the research community

Many of the issues around data access highlighted in Chapter 2 are also issues of inequality, with some researchers having much easier access than others to sufficient quality data. Workshop participants recognised that researchers with access to certain people or institutes were more privileged, leading to "data haves and data have-nots". For instance, those who had access to members of SAGE or who had established relationships with agencies in health and other relevant areas were able to contribute more easily to the research response.

There was also geographical inequality within the data science response, with some areas of the UK having far greater access to local data and success working with it than others. Participants commented that higher quality, spatio-temporal data could have assisted the support of policy at both a local and national level.

Finally, participants noted a lack of diversity in the data science and AI community, which could bias research or policy in a number of ways, and emphasises the need to make diversity more of a priority in recruitment. This lack of diversity could lead to lower engagement with vulnerable and underrepresented groups, for instance, potentially resulting in data and policy biases due to insufficient representation. The low take-up of vaccination in some minority ethnic groups demonstrates the need for communication and engagement by trusted representatives⁵⁰ – and these often differ from the demographics of national representatives.

Suggestions

Reflecting on these challenges, participants made the following suggestions for the data science and AI community. We note that many of the suggestions in Chapter 2 regarding data access and standardisation are also relevant to the data privilege issues highlighted in this chapter.

- **Prioritise understanding the impact of COVID-19 on different ethnic and social groups**, and ensure that these insights are considered in conjunction with other key risk factors (e.g. age, pre-existing health conditions, profession, care home residency). This would help promote inclusion of underserved groups by funders, researchers designing studies, reviewers evaluating focus areas, and teams delivering modelling projects.
- **Address deficiencies in data sources**. For example, the physical, mental health and economic consequences of the pandemic are likely to be severe on some groups, and capturing any missing or incomplete data could be key to understanding the medium- and long-term impacts on underrepresented groups.
- **Develop clear protocols for collecting data on protected characteristics**, including, for example, ethnicity, sex, age and socio-economic status. Better monitoring of this information in studies and clinical trials could help identify critical data gaps.
- **Develop clear protocols for generating anonymised and synthetic data**, so that important demographic information can be included in open datasets (one example is the OpenPseudonymiser⁵¹ application developed at the University of Nottingham). Great strides have been made over the past decade in developing realistic synthetic simulation data (e.g. in 'digital twins' projects). Such work can be leveraged and extended to provide synthetic datasets for modelling potential future pandemics and their impact on minority groups.

50 <https://www.theguardian.com/world/2021/mar/25/clinics-pop-up-in-london-to-help-low-vaccine-take-up>

51 <https://www.openpseudonymiser.org>

- **Make more granular data available at the local level**, and expertise and resources easier to find and deploy, so that local organisations have the autonomy to work with their data (as opposed to the data being held, and decisions made, centrally). This can increase local resilience to health emergencies.
- **Increase engagement with minority and underrepresented communities.** Improve data representativeness, minimise algorithmic bias and develop trust by engaging and consulting with representative groups, such as ‘citizen groups’, at regular intervals. Develop mechanisms for well-funded community involvement in setting research priorities and for the end-to-end participatory design of projects.
- **Increase the diversity of representatives in academia, the government and the media.** Increasing diversity and descriptive representation in organisations – among researchers and UK government representatives alike – can create smarter, more innovative and more inclusive teams and solutions.⁵² For example, diverse research teams will be key to addressing the issue of algorithmic bias. Increasing diversity also sets role models for future generations and thus provides lasting benefits. One way to improve diversity is to make it a core criterion in hiring and appointment decisions.

“Increasing diversity and descriptive representation in organisations – among researchers and UK government representatives alike – can create smarter, more innovative and more inclusive teams and solutions.”

⁵² <https://hbr.org/2016/11/why-diverse-teams-are-smarter> ; <https://www.turing.ac.uk/news/where-are-women-mapping-gender-job-gap-ai>

4. Communication

Although collaboration within the data science and AI community increased during the pandemic, as demonstrated by the examples in Chapter 1, issues of communication were frequently raised by workshop participants, falling into three main categories:

- Communication between experts
- Communication between researchers and policy makers
- Communication between researchers and the public

Communication between experts

Connecting expertise – the right people to the right data to the right problem – was identified by workshop participants as an important bottleneck in working to understand COVID-19, constraining the community’s ability to respond. More collaboration between often disparate groups, such as data asset holders, subject matter experts, researchers and clinicians, could have helped to share expertise, avoid duplication, and improve analyses.

Participants also identified a need for more international collaboration. While data access and exchanges across national borders happened to some degree during the pandemic, such as through the EULAR COVID-19 Database,⁵³ many other efforts remained decentralised, informal and ad hoc. More could be learned by the UK from other countries in similar situations, particularly those with greater experience or capacity in dealing with health emergencies.

Suggestions

Participants made the following suggestions for the data science and AI community:

- **Develop collaborative working relationships** between data scientists and clinicians. Data scientists can provide insights about the collection and storage of data that clinicians may not be aware of, while clinicians can provide valuable insights about the multi-dimensional nature of health and social data collection.

⁵³ https://www.eular.org/eular_covid19_database.cfm

⁵⁴ <https://royalsociety.org/-/media/about-us/international/g-science-statements/G7-data-for-international-health-emergencies-31-03-2021.pdf>

- Make greater strides to **bring transdisciplinary groups together**. As well as research groups, this includes local communities, health and social care providers, local partners, analysts, third sector organisations, private organisations, and industry.
- Speed up the research pipeline by **developing a robust research/analysis/review framework** to reduce delays in publishing ideas while maintaining high standards of quality and transparency.
- **Promote global data sharing** for health emergencies. Data sharing agreements between the UK and other countries take time to organise, and many were previously coordinated and/or financially supported by EU organisations, which will undoubtedly become a more complicated endeavour post-Brexit. Having more flexible regulations for future emergencies will help facilitate global learning as well as research and policy solutions in the UK. The editors note that facilitating safe and equitable data sharing between countries is also recommended by the ‘Data for international health emergencies’ statement published by the science academies of the G7 nations in March 2021.⁵⁴

Communication between researchers and policy makers

The transfer of knowledge from academic research to real-world policy was a crucial aspect of the scientific response to the pandemic. Within the data science and AI community, participants noted that communication with policy makers could be improved in order to better identify policy needs and improve knowledge transfer.

Participants also highlighted a lack of transparency in policy-making. It was difficult to know which studies had ‘cut through’ and been considered by government and advisory groups when making policy interventions, and which data policy makers were using to inform their decisions. Increased transparency would help researchers to understand what scientific approaches and insights are most valuable to

policy makers, and would also give the public a clearer picture of the evidence behind policy decisions, potentially increasing public trust and compliance.

"The transfer of knowledge from academic research to real-world policy was a crucial aspect of the scientific response to the pandemic."

Suggestions

Participants made the following suggestions for the data science and AI community:

- **Build sustainable links** between the data science community and those who are closer to policy-making.
- **Provide training for policy makers** to help them understand the research process, e.g. study design and limitations, and the uncertainty around estimates and findings.
- **Provide training for researchers and public health professionals** to help them better communicate to policy makers the findings, limitations and uncertainties of their studies and models.

Communication between researchers and the public

Throughout the pandemic, data science and AI have been in the public eye as never before. Every day presents a new swathe of statistics about COVID-19, and data scientists have been increasingly called upon to communicate their research to non-specialists. Meanwhile, the public has been able to directly input into COVID-19 research through initiatives such as the COVID Symptom Study app,⁵⁵ and there have been vigorous debates on the ethics of AI and algorithmic bias, especially around the UK GCSE and A-Level grading controversy in summer 2020.⁵⁶

Workshop participants agreed that, although there had been successful examples of public engagement from the community during the pandemic, there were also shortcomings in communication, particularly around the limitations and uncertainties of research. Helping the public to understand the findings and caveats of modelling studies, for example, could enhance trust in the research and increase support for and compliance with policies. Trust and acceptance of policies might also be increased by communicating how data science and modelling can promote positive health outcomes. For example, targeted communication might have helped improve compliance with the isolation notifications sent through the NHS Test and Trace system.⁵⁷

Suggestions

Participants made the following suggestions for the data science and AI community:

- **Provide clear, simplified, accessible information** for non-specialists about studies' findings and predictions, as well as the underlying research designs (data used, data quality, computational processes, etc.).
- **Communicate transparently about the limitations of studies**, such as uncertainty in the models/predictions and potential biases in the data.
- **Build public trust** by addressing concerns related to data security, privacy and confidentiality, and by communicating how the data and models are informing policies.
- **Take a more proactive role in countering misinformation**, for example about testing, contact tracing and vaccines. More effective communication could have mitigated against the many counter-messages, 'fake news' stories, and biased reporting that emerged during the pandemic. Also, present data in a way that is complete and not prone to misinterpretation.
- **Train researchers** in how to communicate their findings to the public and media, especially around the use of models in making decisions and predictions.

- **Increase the understanding and involvement of data scientists in media processes.** Media outlets have increasingly referenced preprints, which are of varying quality given preprint protocols and accelerated peer review processes. Communicate to the public what is and what is not (or not yet) quality, peer-reviewed research.
- **Consider building on pre-existing mechanisms to support communication**, for example via the Science Media Centre,⁵⁸ which curates expert reactions to noteworthy science stories, helping journalists to access nuanced and unfiltered information.

- **Use data visualisation:** this can be an effective means of communicating complex scientific topics to non-specialists. Participants cited Harry Stevens's 'corona simulator' article in *The Washington Post* as a particularly effective example.⁵⁹
- **Support independent input and communication:** workshop participants highlighted the benefits of an Independent SAGE,⁶⁰ which provided a model for how scientists can analyse, interpret and comment on a situation from a more politically neutral standpoint.

⁵⁵ <https://covid.joinzoe.com>

⁵⁶ https://en.wikipedia.org/wiki/2020_UK_GCSE_and_A-Level_grading_controversy

⁵⁷ <https://www.medrxiv.org/content/10.1101/2020.09.15.20191957v1>

⁵⁸ <https://www.sciencemediacentre.org>

⁵⁹ <https://www.washingtonpost.com/graphics/2020/world/corona-simulator>

⁶⁰ <https://www.independentsage.org>

“Throughout the pandemic, data science and AI have been in the public eye as never before.”

5. Conclusions

Since the first UK lockdown in March 2020, when little was known about COVID-19, our knowledge of the disease and its effects have much progressed. The UK's data science and AI community has played an integral role in this scientific effort, and this report provides a snapshot of the community's contributions. But equally, the pandemic has pulled into sharp focus a number of areas where we can and should do better.

First, participants reported difficulties accessing sufficiently timely, robust, granular, standardised and documented data. There were also issues of privilege, with some researchers able to access data much more easily than others.

Second, the pandemic has highlighted pre-existing societal issues of inequality and exclusion, and the role that data science and AI can play in mitigating or exacerbating these. In our workshops, participants raised concerns about a lack of data about, and engagement with, minority ethnic and socio-economic groups, and a lack of diversity within the research community and decision-making organisations.

Third, the pandemic has underlined both the importance and difficulty of communicating transparently with other researchers, policy makers and the public, particularly around issues of modelling and uncertainty. Participants agreed that better communication

is key to reducing the chances of data and research being misused or misinterpreted.

Workshop participants have made many suggestions for how the community might address these challenges. These include the provision of accessible and centralised 'data lakes'; more equitable data access; protocols for data standardisation; increased representation of, and engagement with, minority groups; training for researchers in communicating their work to non-specialists; and initiatives to increase public understanding of research findings and uncertainty, to better counter misinformation.

If the community can make progress in these areas, then when we are next faced with a pandemic – and the historical record strongly suggests that this is a 'when' rather than an 'if' – we should be better placed as a collective to respond.

Navigating our way through the pandemic without the knowledge and resources of the data science and AI community would have been markedly more difficult. These are transformational times for the community as its research becomes ever more embedded in everyday life. We need to draw on our experiences during this pandemic to ensure that data science and AI continue to change lives for the better.

Acknowledgements

We thank all conference and workshop participants who have shared their experiences and insights with us, and have contributed to this report.

We thank Jessie Wand for her amazing efforts in organising the complex series of conference meetings and workshops that led to this report, and James Lloyd, science writer, for his tireless efforts in pulling the final draft of this report together. We are also grateful to the Centre for Facilitation for holding the workshops and writing the workshop notes.

This work was supported by Wave 1 of the UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the 'Health' theme within that grant and The Alan Turing Institute. We also acknowledge the contributions of the 'Health and medical sciences' programme at The Alan Turing Institute.

Appendix A. Workshop participants and Organising Committee

We are pleased to acknowledge the involvement of a great many individuals who participated in our workshops. We have done our best to identify everyone who participated, in whatever form: please accept our heartfelt apologies if we have inadvertently omitted you.

We extend a special thanks to the theme leads, who took on extra responsibilities and were especially generous with their time. Note that the views expressed in this report do not necessarily reflect those of all the individuals named below.

Workshop participants (theme leads in bold)

Adam Bancroft (Bournemouth University), Amitava Banerjee (UCL), Andrea Baronchelli (City, University of London / The Alan Turing Institute), Mark Birkin (The Alan Turing Institute / University of Leeds), **Mark Briers (The Alan Turing Institute)**, Jobie Budd (UCL), Massimo Cavallaro (University of Warwick), **Tao Cheng (UCL)**, Gary Collins (University of Oxford), Louise Coutts (University of Southampton), Helen Curtis (University of Oxford), Dominic Cushnan (NHSX), Alireza Daneshkhah (Coventry University), Leon Danon (University of Exeter), Beatriz de la Iglesia (University of East Anglia), **Spiros Denaxas (UCL)**, **John Dennis (University of Exeter)**, Neil Dhir (The Alan Turing Institute), Karla Diaz-Ordaz (London School of Hygiene & Tropical Medicine), Duygu Dikicioglu (UCL), Rosalind Eggo (London School of Hygiene & Tropical Medicine), Ari Ercole (University of Cambridge), Xiuyi Fan (Swansea University), Daisy Fancourt (UCL), Kate (Katayoun) Farrahi (University of Southampton), Neil Ferguson (Imperial College London), Leo Freitas (Newcastle University), Nicholas Fuggle (University of Southampton / The Alan Turing Institute), Fiona Grimm (The Health Foundation), Prateek Gupta (University of Oxford / The Alan Turing Institute), Chris Hicks (The Alan Turing Institute), Edward Hill (University of Warwick), Alex Holmes

(University of Warwick), Kaveh Jahanshahi (Office for National Statistics), Emily Jarratt (Centre for Data Ethics and Innovation), Mark Jit (London School of Hygiene & Tropical Medicine), Indra Joshi (NHSX), Eiman Kanjo (Nottingham Trent University), Ruth Keogh (London School of Hygiene & Tropical Medicine), Thomas King (RSS Data Ethics SIG), Roman Klapaukh (UCL), Alvina Lai (UCL), Sinead Langan (London School of Hygiene & Tropical Medicine), Saskia Lawson-Tovey (University of Manchester), **Sabina Leonelli (University of Exeter)**, **David Leslie (The Alan Turing Institute)**, Godfrey Leung (Scalpel Ltd.), Myles-Jay Linton (University of Bristol), David Lopez (University of Exeter), Zudi Lu (University of Southampton), Miguel Angel Luque-Fernandez (London School of Hygiene & Tropical Medicine), Samantha Lycett (University of Edinburgh), Ben MacArthur (The Alan Turing Institute / University of Southampton), Maxine Mackintosh (The Alan Turing Institute / The Health Foundation), Varun Maddali (Fintech Sandpit), **Marion Mafham (University of Oxford)**, **Ed Manley (University of Leeds)**, Simon Maskell (University of Liverpool), Kevin McConway (The Open University), Angela McLean (Ministry of Defence), Kelly McLean (PETRA Data Science), Sally McManus (NatCen / City, University of London), Graham Medley

(London School of Hygiene & Tropical Medicine), David Menon (University of Cambridge), Victoria Mico (UKRI EPSRC), Fayyaz Minhas (University of Warwick), Paolo Missier (Newcastle University), **Karyn Morrissey (University of Exeter)**, Louisa Nolan (Data Science Campus, Office for National Statistics), Jamil Nur (University of Cambridge), Lucy Okell (Imperial College London), Jasmina Panovska-Griffiths (UCL), **Deepak Parashar (University of Warwick)**, Imogen Parker (Ada Lovelace Institute), Diego Andres Perez Ruiz (University of Manchester), Arwa Raies (European Bioinformatics Institute), Camila Rangel Smith (The Alan Turing Institute), Francisco Rowe (University of Liverpool), Manu Shankar-Hari (Guy's

and St Thomas' NHS Foundation Trust), Susan Shenkin (University of Edinburgh), Ian Simpson (University of Edinburgh), Tamir Sirkis (University of Exeter), Helen Smith (University of Bristol), Giuseppe Sollazzo (NHSX), Emma Southall (University of Warwick), Isabel Stewart (Haringey Council), Nicola Stingelin-Giles (University of Basel), Nel Swanepoel (UCL), John Tazare (London School of Hygiene & Tropical Medicine), Cagatay Turkey (University of Warwick), Sam Van Stroud (UCL), **Jim Weatherall (AstraZeneca)**, Yorick Wilks (University of Sheffield), Sophie Williams (Barts Health NHS Trust), Arkadiusz Wiśniowski (University of Manchester), Richard Wood (NHS Bristol, North Somerset and South Gloucestershire CCG).

Organising Committee

We also acknowledge the important contributions of the Organising Committee in helping us to shape the format and structure of the conference and workshops.

Miguel Bernabeu (University of Edinburgh), Mark Birkin (The Alan Turing Institute / University of Leeds), Inken von Borzyskowski (UCL), Jon Crowcroft (The Alan Turing Institute / University of Cambridge), Andrew Dowsey (The Alan Turing Institute / University of Bristol), Lilian Edwards (Newcastle University), Ian Hall (The Alan Turing Institute / University of Manchester), Chris Holmes (The Alan Turing Institute / University of Oxford), Frank Kelly (The Alan Turing Institute / University of Cambridge), David Llewellyn (The Alan Turing

Institute / University of Exeter), Ben MacArthur (The Alan Turing Institute / University of Southampton), Eirini Malliaraki (The Alan Turing Institute / University of Oxford), Bilal Mateen (The Alan Turing Institute / Wellcome Trust / King's College NHS), Anjali Mazumder (The Alan Turing Institute), Jon Rowe (The Alan Turing Institute / University of Birmingham), Joanna Stacey (The Alan Turing Institute), Michael Tildesley (University of Warwick), Jessie Wand (The Alan Turing Institute), Adrian Weller (The Alan Turing Institute / University of Cambridge), Darren Wilkinson (The Alan Turing Institute / Newcastle University), Michael Wooldridge (The Alan Turing Institute / University of Oxford).

Appendix B. Workshop themes and reports

The workshops covered the following eight themes, structured around four main areas (theme leads in brackets):

- **Part A: Public health, modelling and pharmaceutical interventions**
 1. Pathogenesis and virus evolution, vaccines and clinical trials (Marion Mafham, Jim Weatherall)
 2. Epidemiological modelling and prediction (Spiros Denaxas, Deepak Parashar)
- **Part B: Non-pharmaceutical interventions**
 3. Testing, contact tracing and other public safety interventions (Mark Briers)
 4. Behavioural analysis and policy interventions (Tao Cheng, Ed Manley)
- **Part C: Impacts**
 5. Non-COVID-related health impacts (Bilal Mateen)
 6. Economic and social impacts (Karyn Morrissey)
- **Part D: Enabling data science response**
 7. Ethics, law and governance (David Leslie)
 8. Data readiness, collection and monitoring (John Dennis, Sabina Leonelli)

The individual workshop reports can be read [here](#).

turing.ac.uk
@turinginst