



# Using imaging to combat a pandemic: rationale for developing the UK National COVID-19 Chest Imaging Database

Joseph Jacob <sup>1,2</sup>, Daniel Alexander<sup>2</sup>, J. Kenneth Baillie <sup>3,4</sup>, Rosalind Berka<sup>5</sup>, Ottavia Bertolli<sup>5</sup>, James Blackwood<sup>6</sup>, Iain Buchan <sup>7</sup>, Claire Bloomfield<sup>8</sup>, Dominic Cushman<sup>9</sup>, Annemarie Docherty<sup>10</sup>, Anthony Edey<sup>11</sup>, Alberto Favaro<sup>5</sup>, Fergus Gleeson<sup>8,12</sup>, Mark Halling-Brown<sup>13,14</sup>, Samanjit Hare<sup>15</sup>, Emily Jefferson<sup>16,17</sup>, Annette Johnstone<sup>18</sup>, Myles Kirby<sup>5</sup>, Ruth McStay<sup>19</sup>, Arjun Nair<sup>20</sup>, Peter J.M. Openshaw<sup>21</sup>, Geoff Parker<sup>2,22</sup>, Gerry Reilly<sup>16</sup>, Graham Robinson<sup>23</sup>, Giles Roditi<sup>24</sup>, Jonathan C.L. Rodrigues<sup>23</sup>, Neil Sebire <sup>16</sup>, Malcolm G. Semple<sup>25</sup>, Catherine Sudlow<sup>26,27</sup>, Nick Woznitza <sup>28,29,30</sup> and Indra Joshi<sup>9</sup>

**Affiliations:** <sup>1</sup>Dept of Respiratory Medicine, University College London, London, UK. <sup>2</sup>Centre for Medical Image Computing, Dept of Computer Science, University College London, London, UK. <sup>3</sup>Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Edinburgh, UK. <sup>4</sup>Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK. <sup>5</sup>Faculty, London, UK. <sup>6</sup>The Industrial Centre for Artificial Intelligence Research in Digital Diagnostics (iCAIRD), Dept of eHealth, NHS Greater Glasgow and Clyde, Glasgow, UK. <sup>7</sup>Institute of Population Health, University of Liverpool, Liverpool, UK. <sup>8</sup>National Consortium of Intelligent Medical Imaging (NCIMI), The University of Oxford, Big Data Institute, Oxford, UK. <sup>9</sup>NHSX, London, UK. <sup>10</sup>Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK. <sup>11</sup>Dept of Radiology, Southmead Hospital, North Bristol NHS Trust, Bristol, UK. <sup>12</sup>Dept of Oncology, University of Oxford, Oxford, UK. <sup>13</sup>Scientific Computing, Royal Surrey NHS Foundation Trust, Guildford, UK. <sup>14</sup>Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, UK. <sup>15</sup>Dept of Radiology, Royal Free London NHS Trust, London, UK. <sup>16</sup>Health Data Research UK, London, UK. <sup>17</sup>Health Informatics Centre (HIC), School of Medicine, University of Dundee, Dundee, UK. <sup>18</sup>Dept of Radiology, Leeds Teaching Hospitals NHS Trust, Leeds General Infirmary, Leeds, UK. <sup>19</sup>Dept of Radiology, Freeman Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. <sup>20</sup>Dept of Radiology, University College London Hospital, London, UK. <sup>21</sup>National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, UK. <sup>22</sup>Bioxydyn Limited, Manchester, UK. <sup>23</sup>Dept of Radiology, Royal United Hospitals Bath NHS Foundation Trust, Bath, UK. <sup>24</sup>Dept of Radiology, University of Glasgow, Glasgow Royal Infirmary, Glasgow, UK. <sup>25</sup>NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK. <sup>26</sup>Usher Institute, University of Edinburgh, Edinburgh, UK. <sup>27</sup>British Heart Foundation (BHF) Data Science Centre, Health Data Research UK, Edinburgh, UK. <sup>28</sup>Radiology Dept, Homerton University Hospital, London, UK. <sup>29</sup>School of Allied and Public Health Professions, Canterbury Christ Church University, Canterbury, UK. <sup>30</sup>12 NHS Nightingale Hospital London, London, UK.

**Correspondence:** Joseph Jacob, Centre for Medical Image Computing, 1st Floor, 90 High Holborn, London, WC1V 6LJ, UK. E-mail: j.jacob@ucl.ac.uk



@ERSpublications

The National COVID-19 Chest Imaging Database (NCCID) is a repository of chest radiographs, CT and MRI images and clinical data from COVID-19 patients across the UK, to support research and development of AI technology and give insight into COVID-19 disease <https://bit.ly/3eQueuha>

**Cite this article as:** Jacob J, Alexander D, Baillie JK, *et al.* Using imaging to combat a pandemic: rationale for developing the UK National COVID-19 Chest Imaging Database. *Eur Respir J* 2020; 56: 2001809 [<https://doi.org/10.1183/13993003.01809-2020>].

Received: 15 May 2020 | Accepted: 8 June 2020

Copyright ©ERS 2020. This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

## Introduction

Since the emergence of the novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), in Wuhan, China in 2019 [1], the resulting coronavirus disease 2019 (COVID-19) disease has rapidly transitioned into a global pandemic with over 31 000 deaths and 215 000 infections in the UK as of 10 May 2020 [2]. Confirmation of SARS-CoV-2 infection requires reverse transcriptase PCR (RT-PCR) testing [3]. Chest radiographic and/or computed tomography (CT) imaging is also central to diagnosis and management [4].

The scale of the COVID-19 pandemic has resulted in the acquisition of huge volumes of imaging data. Traditionally, research using imaging data constituted collation of data within single hospitals or groups of hospitals at most. Endeavours on a local scale have the constraint that not all patient subgroups or disease manifestations might be captured in the collected data. It has long been recognised that there is an acute need to curate larger, more comprehensive datasets to better understand a disease. COVID-19 has arrived in an era where advances in computational power, aligned with an increased availability of big data and the development of self-learning neural networks, have begun to redefine research in medicine. In recent years, computer algorithms trained on imaging data, widely available on the internet, have been adapted to the task of medical image analysis [5, 6]. For computer algorithms to be successfully applied to medical image analysis, it is imperative that they train on large volumes and representative examples of imaging data. These are typically orders of magnitude larger than traditional imaging research datasets, and beyond the capacity of traditional research e-infrastructure.

The National Health Service (NHS) in the UK has long sought national repositories of linked clinical and imaging data, which are essential for applications of artificial intelligence (AI) computing systems in healthcare. Historically, logistical barriers to this have seemed insurmountable and progress has been notoriously slow. Yet, one aspect of the COVID-19 response has been the issue of a notice under Regulation 3(4) of the Health Service Control of Patient Information Regulations 2002 in the UK, which has temporarily eased data sharing restrictions to facilitate COVID-19-specific public health research and scientific collaboration over the course of the emergency.

The British Society of Thoracic Imaging research network began a multicentred COVID-19 imaging study [7], which grew into a partnership with NHSX to create the National COVID-19 Chest Imaging Database (NCCID; <https://nhsx.github.io/covid-chest-imaging-database>). NCCID has put in place mechanisms to collate all chest imaging and pre-specified clinical data from every UK hospital where patients undergo a RT-PCR test for COVID-19. This will include all RT-PCR-positive patients and a representative sample of RT-PCR-negative patients. The study aims to identify information in COVID-19 imaging that may be inconspicuous to the human eye, but which is extractable by computer algorithms. Such buried information may allow the early identification of patients at risk of deterioration, thereby anticipating future intensive care needs. AI algorithms may also demonstrate how comprehensive characterisation of COVID-19 may improve care/outcomes.

NCCID data and image transfer solutions are robust and secure, including those adapted from techniques tried and tested on numerous research studies involving large-scale medical image collection [8]. To maximise efficient resource utilisation in busy hospitals during the course of the pandemic, we are linking our imaging data to the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) UK Clinical Characterisation Protocol (<https://isaric.tghn.org/UK-CCP>), and aim to link to the Intensive Care National Audit and Research Centre (ICNARC; [www.icnarc.org](http://www.icnarc.org)). ISARIC investigators are collating clinical information and biological samples for COVID-19 cases of all ages admitted to hospitals, while ICNARC collates detailed data from adults in the intensive care setting. The study has also been supported by Health Data Research UK as part of its UK response to COVID-19.

An endeavour of this scale, rarely attempted before in the NHS, can only succeed if the radiology and scientific community contribute their time, effort and available data. ISARIC has open source processes and already has established data and material sharing from UK COVID-19 cases [9]. The NCCID initiative will create an open, well-governed database for researchers from academia and industry to add to COVID-19 knowledge collectively. Public, patient and professional trust is vital to NCCID and COVID-19 in general, and we consider accessibility and feedback on data uses as central to good governance. The NCCID data access committee aim to link to researchers who are asking related scientific questions, using complementary methodologies. NCCID will also set aside a portion of its data to enable validation of AI models. This will maintain the highest standards of governance throughout the emergency and allow technology developers to validate their algorithms promptly.

## NCCID data

### Imaging data

#### Chest radiographs

The UK prioritised the use of chest radiographs in the clinical work-up of patients suspected of COVID-19 [10]. The choice of chest radiographs as the primary diagnostic imaging test was pragmatic, given the reported limited sensitivity of CT imaging in COVID-19 diagnosis [11]. The choice also reflected concerns regarding seeding of infection *via* CT scanners contaminated with virus particles, as well as the limited numbers of CT scanners serving the UK population when compared with other European countries [12]. Accordingly, a large proportion of imaging in NCCID will consist of chest radiographs acquired at initial presentation of the patient to hospital and throughout the patient's hospital stay (figure 1). NCCID will collect chest radiographs from all RT-PCR-positive COVID-19 patients in hospitals throughout the UK. In addition, a number of chest radiographs from RT-PCR-negative patients will be collected from sites as a representative control population.

#### CT chest imaging

NCCID will collect all chest CT imaging in RT-PCR-positive COVID-19 patients. This will include non-contrast-enhanced chest CT imaging, CT pulmonary angiograms and CT coronary angiograms.

#### Previous chest imaging

For all RT-PCR-positive COVID-19 patients, NCCID will acquire all chest imaging performed in the previous 3 years. For RT-PCR-negative patients, NCCID will acquire any chest imaging performed in the previous 4 weeks.

### Clinical data

Clinical data collected from RT-PCR-positive COVID-19 patients will include demographic information, patient comorbidity, smoking and medication histories, clinical observations, admission blood test results,

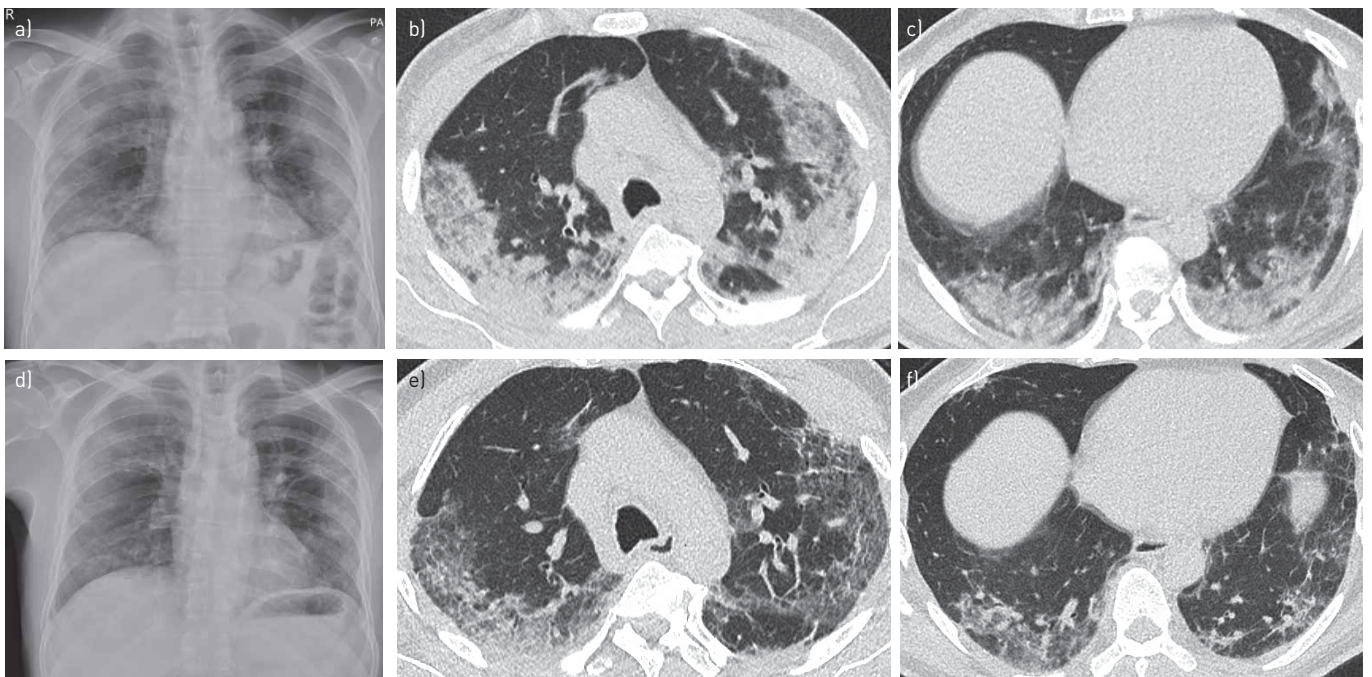


FIGURE 1 Imaging from a 61-year-old man presenting with a 7-day history of feeling unwell and increasing breathlessness. a) The chest radiograph on admission demonstrated patchy peripheral consolidation in the middle and lower zones. Axial CT images of the b) upper and c) lower lung zones on day 6 of admission confirmed multifocal peripheral consolidation, in keeping with the diagnosis of COVID-19. d) A repeat chest radiograph taken 3.5 weeks later showed marked resolution of the consolidation, confirmed on repeat CT imaging of e) upper and f) lower lung zones, although a degree of architectural distortion was then evident within the lungs. Artificial intelligence tools, when applied to imaging and corresponding clinical data at scale, can allow the quantitation of patterns of disease progression in COVID-19. Unsupervised deep learning methods could help identify discrete COVID-19 patient phenotypes that manifest differing disease courses. Image analysis may also allow prediction of those patients that are likely to manifest chronic multisystem sequelae following their acute infection, which may inform future resource prioritisation in healthcare systems.

and outcomes including time of intensive care unit admission and survival. The imaging will link to other clinical databases such as ISARIC to allow analyses against more detailed clinical information.

### **Data access**

We anticipate that, in May 2020, NCCID data will be released to interested academic and commercial groups. Given that NCCID will be of particular interest to AI researchers, a portion of the data will be segregated to allow rigorous and independent validation of AI models. Data access will be initiated through online applications, to be filled in and submitted according to the instructions on the NCCID website. Applications will then be assessed by a central data access committee comprising scientific advisors, technology advisors, information-governance advisors, patient/ethics advisors and system advisors, to evaluate the positive impact on the NHS overall. Successful applicants can access the data on a cloud-based Amazon S3 bucket, and transfer it onto their computing infrastructure, which are required to fulfil high standards of IT security.

### **Discussion**

In these challenging times, research efforts necessary to better understand COVID-19 cannot afford to be fragmented and uncoordinated. Accelerating insights and discovery necessitates leveraging economies of scale and resource amongst researchers, institutions and companies. We believe NCCID will quickly improve COVID-19 understanding and patient care and build on early insights into COVID-19 [13–15].

The scale of the current pandemic requires the best minds, data, algorithms and research programmes to come together quickly. We are fortunate in the UK to have an abundance of internationally recognised researchers working on COVID-19. We hope that NCCID provides a common UK resource to fuel international efforts in tackling COVID-19 and allows better preparedness for similar future needs.

Conflict of interest: J. Jacob reports fees from Boehringer Ingelheim and Roche unrelated to the current submission and is supported by a Clinical Research Career Development Fellowship 209553/Z/17/Z from the Wellcome Trust, and by the NIHR Biomedical Research Centre at University College London. D. Alexander has nothing to disclose. J.K. Baillie reports grants from DHSC National Institute of Health Research UK, Medical Research Council UK, Wellcome Trust, Fiona Elizabeth Agnew Trust, Intensive Care Society and Chief Scientist Office, during the conduct of the study. R. Berka is an Associate at Faculty; Faculty has a paid partnership with NHSX to build its new AI lab to help drive digital transformation and the use of AI in the NHS. This contract has been extended for setting up the platform environment that manages the data storage for the National COVID-19 Chest Imaging Database (NCCID); the platform is being made available by Faculty at zero licence cost. In addition, Faculty is contracted to support NHS England and Improvement as well as NHSX with its data response strategy to COVID-19, which includes developing dashboards, models and simulations to provide information to central government decision-makers. O. Bertolli is an Associate at Faculty; Faculty has a paid partnership with NHSX to build its new AI lab to help drive digital transformation and the use of AI in the NHS. This contract has been extended for setting up the platform environment that manages the data storage for the National COVID-19 Chest Imaging Database (NCCID); the platform is being made available by Faculty at zero licence cost. In addition, Faculty is contracted to support NHS England and Improvement as well as NHSX with its data response strategy to COVID-19, which includes developing dashboards, models and simulations to provide information to central government decision-makers. J. Blackwood has nothing to disclose. I. Buchan reports personal fees from and acting as advisor for AstraZeneca, grants from NIHR, and is a former employee of Microsoft, outside the submitted work. C. Bloomfield reports that NCIMI is funded through support from the Industry Strategy Challenge Fund, by Innovate UK grant 104688. D. Cushman has nothing to disclose. A. Docherty has nothing to disclose. A. Edey has nothing to disclose. A. Favaro is a Lead Data Scientist at Faculty; Faculty has a paid partnership with NHSX to build its new AI lab to help drive digital transformation and the use of AI in the NHS. This contract has been extended for setting up the platform environment that manages the data storage for the National COVID-19 Chest Imaging Database (NCCID); the platform is being made available by Faculty at zero licence cost. In addition, Faculty is contracted to support NHS England and Improvement as well as NHSX with its data response strategy to COVID-19, which includes developing dashboards, models and simulations to provide information to central government decision-makers. F. Gleeson has nothing to disclose. M. Halling-Brown is an advisor to Google Health on the MAMMOTH project, and has a Visiting Professorship with University of Surrey as part of a research collaboration with Transpara. S. Hare has nothing to disclose. E. Jefferson reports grants from Medical Research Council (MRC), Health Data Research (HDR) UK, National Institute of Health Research (NIHR), Chief Scientist Office (CSO), Engineering and Physical Sciences Research Council (EPSRC), Health Foundation, Data Lab, Scottish Government, NHS Fife Health Board and EU Horizon 2020, during the conduct of the study. A. Johnstone has nothing to disclose. M. Kirby is a Principal at Faculty; Faculty has a paid partnership with NHSX to build its new AI lab to help drive digital transformation and the use of AI in the NHS. This contract has been extended for setting up the platform environment that manages the data storage for the National COVID-19 Chest Imaging Database (NCCID); the platform is being made available by Faculty at zero licence cost. In addition, Faculty is contracted to support NHS England and Improvement as well as NHSX with its data response strategy to COVID-19, which includes developing dashboards, models and simulations to provide information to central government decision-makers. R. McStay has nothing to disclose. A. Nair reports salary reimbursement from Biomedical Research Centre UCL, and is medical advisor to Aidence BV, outside the submitted work. P.J.M. Openshaw reports personal fees for consultancy work with Janssen, J and J and Sanofi; grants from MRC, European Union, NIHR Biomedical Research Centre and Wellcome Trust, collaborative grants with GSK, personal fees from the European Respiratory Society and an NIHR Senior Investigator Award outside the submitted work; in addition, P.J.M. Openshaw was President of the British Society for Immunology; this was an unpaid appointment but travel and accommodation at

some meetings was provided by the Society. G. Parker is director and shareholder of Bioxydyn Limited, outside the submitted work. G. Reilly has nothing to disclose. G. Robinson has nothing to disclose. G. Roditi has nothing to disclose. J.C.L. Rodrigues has nothing to disclose. N. Sebire has nothing to disclose. M.G. Semple reports grants from DHSC National Institute of Health Research UK, Medical Research Council UK, Health Protection Research Unit in Emerging and Zoonotic Infections, and University of Liverpool, during the conduct of the study; and is minority owner of Integrum Scientific LLC, Greensboro, NC, USA, outside the submitted work. C. Sudlow has nothing to disclose. N. Woznitza reports grants from Cancer Research UK and Roy Castle Lung Cancer Foundation, personal fees from InHealth, outside the submitted work. I. Joshi has nothing to disclose.

## References

- 1 Huang C, Wang Y, Li X, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395: 497–506.
- 2 Johns Hopkins University Center for Systems Science and Engineering. COVID-19 Dashboard: World Map. <https://coronavirus.jhu.edu/map.html>
- 3 Guan WJ, Chen RC, Zhong NS. Strategies for the prevention and management of coronavirus disease 2019. *Eur Respir J* 2020; 55: 2000597.
- 4 Rodrigues JCL, Hare SS, Edey A, *et al.* An update on COVID-19 for the radiologist – a British Society of Thoracic Imaging statement. *Clin Radiol* 2020; 75: 323–325.
- 5 De Fauw J, Ledsam JR, Romera-Paredes B, *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018; 24: 1342–1350.
- 6 Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115–118.
- 7 Hare SS, Rodrigues JCL, Jacob J, *et al.* A UK-wide British Society of Thoracic Imaging COVID-19 imaging repository and database: design, rationale and implications for education and research. *Clin Radiol* 2020; 75: 326–328.
- 8 Halling-Brown MD, Warren LM, Ward D, *et al.* OPTIMAM Mammography Image Database: a large scale resource of mammography images and clinical data. *arXiv* 2020; preprint [<https://arxiv.org/abs/2004.04742>].
- 9 Dunning JW, Merson L, Rohde GGU, *et al.* Open source clinical science for emerging infections. *Lancet Infect Dis* 2014; 14: 8–9.
- 10 Nair A, Rodrigues JCL, Hare S, *et al.* A British Society of Thoracic Imaging statement: considerations in designing local imaging diagnostic algorithms for the COVID-19 pandemic. *Clin Radiol* 2020; 75: 329–334.
- 11 Bernheim A, Mei X, Huang M, *et al.* Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology*; 295: 200463.
- 12 EUROSTAT. Healthcare resource statistics – technical resources and medical technology. [https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare\\_resource\\_statistics\\_-\\_technical\\_resources\\_and\\_medical\\_technology](https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_resource_statistics_-_technical_resources_and_medical_technology) Date last updated: 28 November 2019.
- 13 Yang S, Shi Y, Lu H, *et al.* Clinical and CT features of early stage patients with COVID-19: a retrospective analysis of imported cases in Shanghai, China. *Eur Respir J* 2020; 55: 2000407.
- 14 Zhang S, Li H, Huang S, *et al.* High-resolution computed tomography features of 17 cases of coronavirus disease 2019 in Sichuan province, China. *Eur Respir J* 2020; 55: 2000334.
- 15 Wang L, Gao YH, Lou LL, *et al.* The clinical dynamics of 18 cases of COVID-19 outside of Wuhan, China. *Eur Respir J* 2020; 55: 2000398.