Early View

Research highlights

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

Joseph Jacob, Daniel Alexander, J. Kenneth Baillie, Rosalind Berka, Ottavia Bertolli, James Blackwood, Iain Buchan, Claire Bloomfield, Dominic Cushnan, Annemarie Docherty, Anthony Edey, Alberto Favaro, Fergus Gleeson, Mark Halling-Brown, Samanjit Hare, Emily Jefferson, Annette Johnstone, Myles Kirby, Ruth Mcstay, Arjun Nair, Peter J.M. Openshaw, Geoff Parker, Gerry Reilly, Graham Robinson, Giles Roditi, Jonathan C.L. Rodrigues, Neil Sebire, Malcolm G. Semple, Catherine Sudlow, Nick Woznitza, Indra Joshi

RESEARCH HIGHLIGHT

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

Corresponding author: Dr Joseph Jacob, Centre for Medical Image Computing, 1st Floor, 90 High Holborn, London, UK, WC1V 6LJ

Email address: j.jacob@ucl.ac.uk

Word Count: 1261

Affiliations

[1] Department of Respiratory Medicine, University College London, London, WC1E 6JF

[2] Centre for Medical Image Computing, Department of Computer Science, University College London, Gower Street, London, WC1E 6BT

[3] Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Edinburgh, UK

[4] Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK

[5] Faculty, 54 Welbeck Street, London W1G 9XS, United Kingdom

[6] iCAIRD, Department of eHealth, NHS Greater Glasgow & Clyde, Scotland, UK.

[7] Institute of Population Health, University of Liverpool, Liverpool, UK.

[8] National Consortium of Intelligent Medical Imaging (NCIMI), The University of Oxford, Big Data Institute, Old Road Campus, Roosevelt Drive, Oxford, OX3 7LF

[9] NHSX, Skipton House, 80 London Road, London, SE1 6LH, UK

[10] Centre for Medical Informatics, Usher Institute, University of Edinburgh, EH16 4SA

[11] Department of Radiology, Southmead Hospital, North Bristol NHS Trust, Southmead Road, Bristol, BS10 5NB

[12] Department of Oncology, University of Oxford, UK

[13] Scientific Computing, Royal Surrey NHS Foundation Trust, Egerton Road, Guildford, GU2 7XX

[14] CVSSP, University of Surrey, University of Surrey, Guildford, GU2 7XH, UK

[15] Department of Radiology, Royal Free London NHS Trust, London, Pond Street, London, NW3 2QJ

[16] Health Data Research UK, Gibbs Building, 215 Euston Road, London, NW1 2BE

[17] Health Informatics Centre (HIC), School of Medicine, University of Dundee, DD1 4HN

[18] Department of Radiology, Leeds Teaching Hospitals NHS Trust, Leeds General Infirmary, Great George Street, Leeds, LS1 3EX

[19] Department of Radiology, Freeman Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Freeman Road, Newcastle upon Tyne, NE7 7DN

[20] Department of Radiology, University College London Hospital, 235 Euston Road, London NW1 2BU

[21] National Heart and Lung Institute, Faculty of Medicine, Imperial College London

[22] Bioxydyn Limited, Manchester M15 6SZ

[23] Department of Radiology, Royal United Hospitals Bath NHS Foundation Trust, Combe Park, Bath, BA1 3NG

[24] Department of Radiology, University of Glasgow, Glasgow Royal Infirmary, 16 Alexandra Parade, Glasgow G31 2ER, UK

[25] NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Faculty of Health and Life Sciences, University of Liverpool.

[26] Usher Institute, University of Edinburgh

[27] BHF Data Science Centre, Health Data Research UK

[28] Radiology Department Homerton University Hospital, Homerton Row London E9 6SR

[29] School of Allied and Pubic Health Professions, Canterbury Christ Church University North Holmes Road, Canterbury CT1 1QU

[30] 12 NHS Nightingale Hospital London, Royal Victoria Dock, 1 Western Gateway, Royal Docks, London E16 1XL

Declarations / Conflicts of interest

INTRODUCTION

Since the emergence of the novel coronavirus SARS-CoV-2 in Wuhan, China in 2019(1), the resulting COVID-19 disease has rapidly transitioned into a global pandemic with over 31,000 deaths and 215,000 infections in the UK as of 10th May 2020(2). Confirmation of SARS-CoV-2 infection requires reverse-transcriptase polymerase chain reaction (RT-PCR) testing(3). Chest radiographic and/or computed tomography 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 sub-groups 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 in the United Kingdom 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

(COPI notice) 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(7) began a multicentred COVID-19 imaging study which grew into a partnership with NHSX to create the National COVID-19 Chest Imaging Database (NCCID)(8). 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.

The NCCID data and image transfer solutions are robust and secure, including those having been adapted from techniques tried and tested on numerous research studies involving large-scale medical image collection(9). To maximise efficient resource utilisation in busy hospitals during the course of the pandemic, we are linking our imaging data to the ISARIC WHO Clinical Characterisation Protocol for Severe Emerging Infection UK (ISARIC CCP-UK)(10) and aim to link to ICNARC(11). ISARIC investigators are collating clinical information and biological samples for COVID-19 cases of all ages admitted to hospitals, whilst 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(12). 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

1) Chest Radiographs: The United Kingdom prioritised the use of chest radiographs in the clinical work-up of patients suspected of COVID-19(13). 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(14). 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(15). 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 patients hospital stay (Figure 1). NCCID will collect chest radiographs in all RT-PCR COVID-19 positive patients in hospitals throughout the UK. In addition, a number of chest radiographs from RT-PCR COVID-19 negative patients will be collected from sites as a representative control population.

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

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

Clinical data

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

DISCUSSION

We anticipate that in May 2020 the NCCID data will be released to interested academic and commercial groups. Given that the 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 website of NCCID (8). 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 on to their computing infrastructure, required to fulfil high standards of IT security.

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(16-18).

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.

REFERENCES

1.	Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020;395(10223):497-506.
2.	Hopkins J. [Available from: https://coronavirus.jhu.edu/map.html.
3.	Guan W-j, Chen R-c, Zhong N-s. Strategies for the prevention and management of coronavirus disease 2019. Eur Respir J. 2020;55(4):2000597.
4.	Rodrigues JCL, Hare SS, Edey A, Devaraj A, Jacob J, Johnstone A, et al. An update on COVID-19 for the radiologist - A British society of Thoracic Imaging statement. Clin Radiol. 2020;75(5):323-5.
5.	De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342-50.
6.	Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-8.
7.	Hare SS, Rodrigues JCL, Jacob J, Edey A, Devaraj A, Johnstone A, 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(5):326-8.
8.	NCCID. 2020 [Available from: https://nhsx.github.io/covid-chest-imaging-database/.
9.	Halling-Brown MD, Warren LM, Ward D, Lewis E, Mackenzie A, Wallis MG, et al. OPTIMAM Mammography Image Database: a large scale resource of mammography images and clinical data. arXiv. 2020.
10.	Isaric Working Group IC. 2019 [Available from: https://isaric.tghn.org/UK-CCP/.
11.	ICNARC. 2019 [Available from: https://www.icnarc.org/.
12.	Dunning JW, Merson L, Rohde GGU, Gao Z, Semple MG, Tran D, et al. Open source clinical science for emerging infections. The Lancet Infectious diseases. 2014;14(1):8-9.
13.	Nair A, Rodrigues JCL, Hare S, Edey A, Devaraj A, Jacob J, 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(5):329-34.
14.	Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. Radiology.0(0):200463.
15.	EUROSTAT. Healthcare resource statistics - technical resources and medical technology. 2019.
16.	Yang S, Shi Y, Lu H, Xu J, Li F, Qian Z, 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(4):2000407.
17.	Zhang S, Li H, Huang S, You W, Sun H. High-resolution computed tomography features of 17 cases of coronavirus disease 2019 in Sichuan province, China. Eur Respir J. 2020;55(4):2000334.
18.	Wang L, Gao Y-H, Lou L-L, Zhang G-J. The clinical dynamics of 18 cases of COVID-19 outside of Wuhan, China. Eur Respir J. 2020;55(4):2000398.
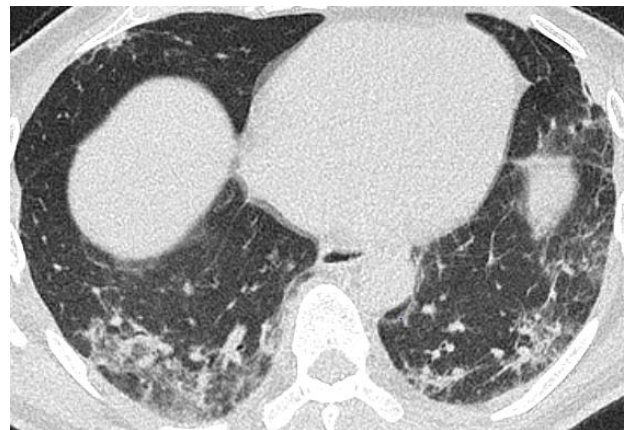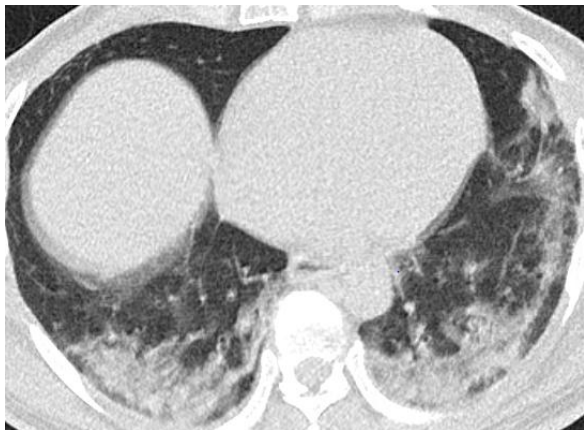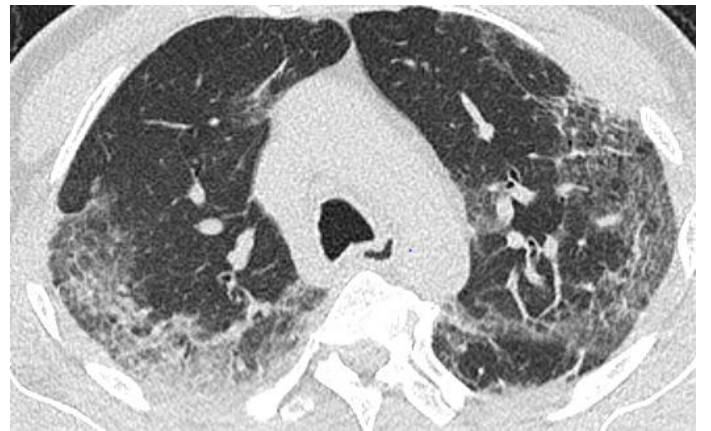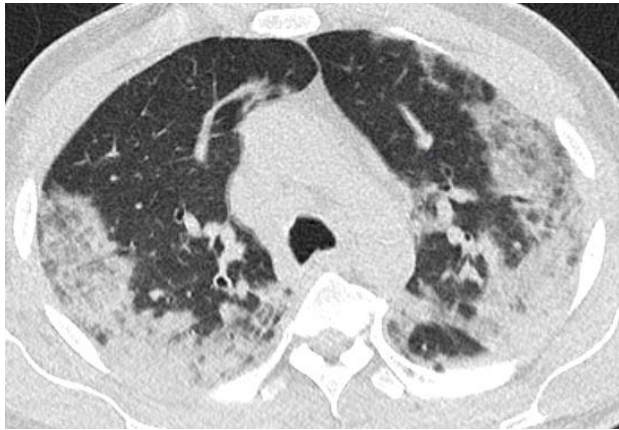
Figure 1. Imaging from a 61-year-old man presenting with a 7-day history of feeling unwell and increasing breathlessness. The left chest radiograph on admission demonstrates patchy peripheral consolidation in the middle and lower zones. Axial CT images of the upper and lower lung zones on day 6 of admission confirms multi-focal peripheral consolidation in keeping with the diagnosis of COVID-19. A repeat chest radiograph taken 3.5 weeks later shows marked resolution of the consolidation, confirmed on repeat CT imaging, though a degree of architectural distortion is now 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.