THE LANCET
Microbe

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Aggarwal D, Myers R, Hamilton WL, et al. The role of viral genomics in understanding COVID-19 outbreaks in long-term care facilities. *Lancet Microbe* 2021; published online Sept 29. https://doi.org/10.1016/S2666-5247(21)00208-1.

Funding acquisition, leadership, supervision, metadata curation, project administration, samples, logistics, Sequencing, analysis, and Software and analysis tools:
Dr Thomas R Connor PhD[33, 34], and Professor Nicholas J Loman PhD[15].

Leadership, supervision, sequencing, analysis, funding acquisition, metadata curation, project administration, samples, logistics, and visualisation:
Dr Samuel C Robson Ph.D [68].

Leadership, supervision, project administration, visualisation, samples, logistics, metadata curation and software and analysis tools:
Dr Tanya Golubchik PhD [27].

Leadership, supervision, metadata curation, project administration, samples, logistics sequencing and analysis:
Dr M. Estee Torok FRCP [8, 10].

Project administration, metadata curation, samples, logistics, sequencing, analysis, and software and analysis tools:
Dr William L Hamilton PhD [8, 10].

Leadership, supervision, samples logistics, project administration, funding acquisition sequencing and analysis:
Dr David Bonsall PhD [27].

Leadership and supervision, sequencing, analysis, funding acquisition, visualisation and software and analysis tools:
Dr Ali R Awan PhD [74].

Leadership and supervision, funding acquisition, sequencing, analysis, metadata curation, samples and logistics:
Dr Sally Corden PhD [33] .

Leadership supervision, sequencing analysis, samples, logistics, and metadata curation:
Professor Ian Goodfellow PhD [11].

Leadership, supervision, sequencing, analysis, samples, logistics, and Project administration:
Professor Darren L Smith PhD [60, 61].

Project administration, metadata curation, samples, logistics, sequencing and analysis:
Dr Martin D Curran PhD [14], and Dr Surendra Parmar PhD [14]**.**

Samples, logistics, metadata curation, project administration sequencing and analysis:
Dr James G Shepherd MBChB MRCP [21].

Sequencing, analysis, project administration, metadata curation and software and analysis tools:
Dr Matthew D Parker PhD [38] and Dr Dinesh Aggarwal MRCP [1, 2, 3].

Leadership, supervision, funding acquisition, samples, logistics, and metadata curation:
Dr Catherine Moore[33] .

Leadership, supervision, metadata curation, samples, logistics, sequencing and analysis:
Dr Derek J Fairley PhD [6, 88], Professor Matthew W Loose PhD [54], and Joanne Watkins MSc [33].

Metadata curation, sequencing, analysis, leadership, supervision and software and analysis tools:
Dr Matthew Bull PhD[33] , and Dr Sam Nicholls PhD [15] .

Leadership, supervision, visualisation, sequencing, analysis and software and analysis tools:
Professor David M Aanensen PhD [1, 30].

Sequencing, analysis, samples, logistics, metadata curation, and visualisation:
Dr Sharon Glaysher [70] .

Metadata curation, sequencing, analysis, visualisation, software and analysis tools:
Dr Matthew Bashton PhD [60], and Dr Nicole Pacchiarini PhD [33].

Sequencing, analysis, visualisation, metadata curation, and software and analysis tools:
Dr Anthony P Underwood PhD [1, 30].

Funding acquisition, leadership, supervision and project administration:
Dr Thushan I de Silva PhD [38], and Dr Dennis Wang PhD [38]**.**

Project administration, samples, logistics, leadership and supervision:
Dr Monique Andersson PhD [28] , Professor Anoop J Chauhan [70], Dr Mariateresa de Cesare PhD[26], Dr Catherine Ludden [1,3] , and Dr Tabitha W Mahungu FRCPath [91].

Sequencing, analysis, project administration and metadata curation:
Dr Rebecca Dewar PhD [20], and Martin P McHugh MSc[20].

Samples, logistics, metadata curation and project administration:
Dr Natasha G Jesudason MBChB MRCP FRCPath [21], Dr Kathy K Li MBBCh FRCPath [21], Dr Rajiv N Shah BMBS MRCP MSc [21], and Dr Yusri Taha MD, PhD [66].

Leadership, supervision, funding acquisition and metadata curation:
Dr Kate E Templeton PhD [20]**.**
Leadership, supervision, funding acquisition, sequencing and analysis:
Dr Simon Cottrell PhD [33], Dr Justin O'Grady PhD [51], Professor Andrew Rambaut DPhil [19], and Professor Colin P Smith PhD[93].

Leadership, supervision, metadata curation, sequencing and analysis:
Professor Matthew T.G. Holden PhD [87], and Professor Emma C Thomson PhD/FRCP [21].

Leadership, supervision, samples, logistics and metadata curation:
Dr Samuel Moses MD [81, 82].

Sequencing, analysis, leadership, supervision, samples and logistics:
Dr Meera Chand [7], Dr Chrystala Constantinidou PhD [71], Professor Alistair C Darby PhD [46], Professor Julian A Hiscox PhD [46], Professor Steve Paterson PhD [46], and Dr Meera Unnikrishnan PhD [71]**.**

Sequencing, analysis, leadership and supervision and software and analysis tools:
Dr Andrew J Page PhD [51], and Dr Erik M Volz PhD[96].

Samples, logistics, sequencing, analysis and metadata curation:
Dr Charlotte J Houldcroft PhD [8], Dr Aminu S Jahun PhD [11], Dr James P McKenna PhD [88], Dr Luke W Meredith PhD [11], Dr Andrew Nelson PhD [61], Sarojini Pandey MSc [72], and Dr Gregory R Young PhD [60].

Sequencing, analysis, metadata curation, and software and analysis tools:
Dr Anna Price PhD[34], Dr Sara Rey PhD [33], Dr Sunando Roy PhD [41], Dr Ben Temperton Ph.D [49], and Matthew Wyles [38].

Sequencing, analysis, metadata curation and visualisation:
Stefan Rooke MSc[19], and Dr Sharif Shaaban PhD[87].

Visualisation, sequencing, analysis and software and analysis tools:
Dr Helen Adams PhD [35], Dr Yann Bourgeois Ph.D [69], Dr Katie F Loveson Ph.D [68], Áine O'Toole MSc[19], and Richard Stark MSc [71].

Project administration, leadership and supervision:
Dr Ewan M Harrison PhD [1, 3], David Heyburn [33], and Professor Sharon J Peacock [2, 3]

Project administration and funding acquisition:
Dr David Buck PhD[26] , and Michaela John BSc Hons [36]

Sequencing, analysis and project administration:
Dorota Jamrozy [1], and Dr Joshua Quick PhD [15]

Samples, logistics, and project administration:
Dr Rahul Batra MD[78], Katherine L Bellis BSc (Hons) [1, 3], Beth Blane BSc [3] , Sophia T Girgis MSc [3], Dr Angie Green PhD [26], Anita Justice MSc [28] , Dr Mark Kristiansen PhD [41] , and Dr Rachel J Williams PhD [41].

Project administration, software and analysis tools:
Radoslaw Poplawski BSc [15].

Project administration and visualisation:
Dr Garry P Scarlett Ph.D[69].

Leadership, supervision, and funding acquisition:
Professor John A Todd PhD [26] , Dr Christophe Fraser PhD [27], Professor Judith Breuer MD [40,41], Professor Sergi Castellano PhD [41], Dr Stephen L Michell PhD [49], Professor Dimitris Gramatopoulos PhD, FRCPath[73], and Dr Jonathan Edgeworth PhD, FRCPath [78].

Leadership, supervision and metadata curation:
Dr Gemma L Kay PhD [51].

Leadership, supervision, sequencing and analysis:
Dr Ana da Silva Filipe PhD [21] , Dr Aaron R Jeffries PhD [49], Dr Sascha Ott PhD [71], Professor Oliver Pybus [24] , Professor David L Robertson PhD [21], Dr David A Simpson PhD [6] , and Dr Chris Williams MB BS[33].

Samples, logistics, leadership and supervision:
Dr Cressida Auckland FRCPath [50], Dr John Boyes MBChB[83], Dr Samir Dervisevic FRCPath[52] , Professor Sian Ellard FRCPath[49, 50] , Dr Sonia Goncalves[1], Dr Emma J Meader FRCPath [51], Dr Peter Muir PhD[2], Dr Husam Osman PhD [95], Reenesh Prakash MPH[52], Dr Venkat Sivaprakasam PhD[18], and Dr Ian B Vipond PhD[2].

Leadership, supervision and visualisation
Dr Jane AH Masoli MBChB [49, 50].

Sequencing, analysis and metadata curation
Dr Nabil-Fareed Alikhan PhD [51], Matthew Carlile BSc [54], Dr Noel Craine DPhil [33], Dr Sam T Haldenby PhD [46], Dr Nadine Holmes PhD [54], Professor Ronan A Lyons MD [37], Dr Christopher Moore PhD [54], Malorie Perry MSc [33] , Dr Ben Warne MRCP[80], and Dr Thomas Williams MD [19].

Samples, logistics and metadata curation:
Dr Lisa Berry PhD [72], Dr Andrew Bosworth PhD [95] ,Dr Julianne Rose Brown PhD[40], Sharon Campbell MSc [67], Dr Anna Casey PhD[17], Dr Gemma Clark PhD [56], Jennifer Collins BSc [66], Dr Alison Cox PhD [43, 44] , Thomas Davis MSc [84], Gary Eltringham BSc [66], Dr Cariad Evans [38, 39] , Dr Clive Graham MD [64], Dr Fenella Halstead PhD [18], Dr Kathryn Ann Harris PhD [40], Dr Christopher Holmes PhD [58], Stephanie Hutchings [2] , Professor Miren Iturriza-Gomara PhD [46] , Dr Kate Johnson [38, 39], Katie Jones MSc [72], Dr Alexander J Keeley MRCP [38], Dr Bridget A Knight PhD [49, 50] , Cherian Koshy MSc, CSci, FIBMS [90], Steven Liggett [63] , Hannah Lowe MSc [81] , Dr Anita O Lucaci PhD [46] , Dr Jessica Lynch PhD MBChB [25, 29] , Dr Patrick C McClure PhD [55] , Dr Nathan Moore MBChB [31] , Matilde Mori BSc [25, 29, 32] , Dr David G Partridge FRCP, FRCPath [38, 39] , Pinglawathee Madona [43, 44] , Hannah M Pymont MSc [2] , Dr Paul Anthony Randell MBBCh [43, 44] , Dr Mohammad Raza [38, 39] , Felicity Ryan MSc [81] , Dr Robert Shaw FRCPath [28], Dr Tim J Sloan PhD [57] , and Emma Swindells BSc [65] .

Sequencing, analysis, Samples and logistics:
Alexander Adams BSc [33], Dr Hibo Asad PhD [33], Alec Birchley MSc [33] , Tony Thomas Brooks BSc (Hons) [41], Dr Giselda Bucca PhD [93], Ethan Butcher [70], Dr Sarah L Caddy PhD [13], Dr Laura G

Caller PhD [2, 3, 12] , Yasmin Chaudhry BSc [11], Jason Coombes BSc (HONS) [33], Michelle Cronin [33], Patricia L Dyal MPhil [41], Johnathan M Evans MSc [33],Laia Fina [33], Bree Gatica-Wilcox MPhil [33], Dr Iliana Georgana PhD [11], Lauren Gilbert A-Levels [33] , Lee Graham BSc [33], Danielle C Groves BA [38], Grant Hall BSc [11], Ember Hilvers MPH [33] , Dr Myra Hosmillo PhD [11], Hannah Jones [33], Sophie Jones MSc [33], Fahad A Khokhar BSc [13] , Sara Kumziene-Summerhayes MSc [33], George MacIntyre-Cockett BSc [26], Dr Rocio T Martinez Nunez PhD [94] , Dr Caoimhe McKerr PhD [33] , Dr Claire McMurray PhD [15], Dr Richard Myers [7], Yasmin Nicole Panchbhaya BSc [41] , Malte L Pinckert MPhil [11] , Amy Plimmer [33] , Dr Joanne Stockton PhD [15] , Sarah Taylor [33] , Dr Alicia Thornton [7] , Amy Trebes MSc [26] , Alexander J Trotter MRes [51] ,Helena Jane Tutill BSc [41] ,Charlotte A Williams BSc [41] , Anna Yakovleva BSc [11] and Dr Wen C Yew PhD [62].

Sequencing, analysis and software and analysis tools:
Dr Mohammad T Alam PhD [71], Dr Laura Baxter PhD [71], Olivia Boyd MSc [96] , Dr Fabricia F. Nascimento PhD [96], Timothy M Freeman MPhil [38], Lily Geidelberg MSc [96], Dr Joseph Hughes PhD [21], David Jorgensen MSc [96], Dr Benjamin B Lindsey MRCP [38], Dr Richard J Orton PhD [21] , Dr Manon Ragonnet-Cronin PhD [96] Joel Southgate MSc [33, 34,] and Dr Sreenu Vattipally PhD [21].

Samples, logistics and software and analysis tools:
Dr Igor Starinskij MSc MRCP [23].

Visualisation and software and analysis tools:
Dr Joshua B Singer PhD [21] , Dr Khalil Abudahab PhD [1, 30], Leonardo de Oliveira Martins PhD[51] , Dr Thanh Le-Viet PhD [51] ,Mirko Menegazzo [30] ,Ben EW Taylor Meng [1, 30], and Dr Corin A Yeats PhD [30].

Project Administration:
Sophie Palmer [3], Carol M Churcher [3] , Dr Alisha Davies [33], Elen De Lacy MSc [33], Fatima Downing [33], Sue Edwards [33] , Dr Nikki Smith PhD [38] , Dr Francesc Coll PhD[97] , Dr Nazreen F Hadjirin PhD[3] and Dr Frances Bolt PhD [44, 45] .

Leadership and supervision:
Dr. Alex Alderton [1], Dr Matt Berriman [1], Ian G Charles [51], Dr Nicholas Cortes MBChB [31] ,Dr Tanya Curran PhD [88] , Prof John Danesh [1], Dr Sahar Eldirdiri MBBS, MSC FRCPath [84], Dr Ngozi Elumogo FRCPath [52], Prof Andrew Hattersley FRS [49, 50], Professor Alison Holmes MD [44, 45], Dr Robin Howe [33], Dr Rachel Jones [33] , Anita Kenyon MSc [84], Prof Robert A Kingsley PhD [51], Professor Dominic Kwiatkowski [1, 9], Dr Cordelia Langford[1], Dr Jenifer Mason MBBS [48], Dr Alison E Mather PhD [51], Lizzie Meadows MA [51], Dr Sian Morgan FRCPath [36], Dr James Price PhD [44, 45], Trevor I Robinson MSc [48] , Dr Giri Shankar [33] , John Wain [51], and Dr Mark A Webber PhD[51] .

Metadata curation:
Dr Declan T Bradley PhD [5, 6] ,Dr Michael R Chapman PhD [1, 3, 4] , Dr Derrick Crooke [28] , Dr David Eyre PhD [28], Professor Martyn Guest PhD[34] , Huw Gulliver [34] , Dr Sarah Hoosdally [28] , Dr Christine Kitchen PhD [34] , Dr Ian Merrick PhD [34], Siddharth Mookerjee MPH [44, 45] , Robert Munn BSc [34] , Professor Timothy Peto PhD[28], Will Potter [52], Dr Dheeraj K Sethi MBBS [52], Wendy Smith [56] ,Dr Luke B Snell MB BS [75, 94] , Dr Rachael Stanley PhD [52] , Claire Stuart [52] and Dr Elizabeth Wastenge MD [20].

Sequencing and analysis:

Dr Erwan Acheson PhD [6] , Safiah Afifi BSc [36] , Dr Elias Allara MD PhD [2, 3] , Dr Roberto Amato [1], Dr Adrienn Angyal PhD[38], Dr Elihu Aranday-Cortes PhD/DVM[21] , Cristina Ariani [1], Jordan Ashworth [19], Dr Stephen Attwood [24], Alp Aydin MSci [51] , David J Baker BEng [51] , Dr Carlos E Balcazar PhD [19], Angela Beckett MSc [68] Robert Beer BSc [36] , Dr Gilberto Betancor PhD[76], Emma Betteridge [1] , Dr David Bibby [7] , Dr Daniel Bradshaw [7] , Catherine Bresner Bsc(Hons) [34], Dr Hannah E Bridgewater PhD [71] , Alice Broos BSc (Hons) [21] , Dr Rebecca Brown PhD [38] , Dr Paul E Brown PhD [71], Dr Kirstyn Brunker PhD [22] , Dr Stephen N Carmichael PhD [21] , Jeffrey K. J. Cheng MSc [71], Dr Rachel Colquhoun DPhil [19] , Dr Gavin Dabrera [7] , Dr Johnny Debebe PhD [54], Eleanor Drury [1], Dr Louis du Plessis [24] , Richard Eccles MSc [46], Dr Nicholas Ellaby [7], Audrey Farbos MSc [49], Ben Farr [1] ,Dr Jacqueline Findlay PhD [41] , Chloe L Fisher MSc [74], Leysa Marie Forrest MSc [41], Dr Sarah Francois [24], Lucy R. Frost BSc [71], William Fuller BSc [34] , Dr Eileen Gallagher [7] , Dr Michael D Gallagher PhD[19] , Matthew Gemmell MSc [46], Dr Rachel AJ Gilroy PhD [51], Scott Goodwin [1], Dr Luke R Green PhD [38] ,Dr Richard Gregory PhD [46] ,Dr Natalie Groves [7] ,Dr James W Harrison PhD [49], Hassan Hartman [7] , Dr Andrew R Hesketh PhD[93],Verity Hill [19], Dr Jonathan Hubb [7] ,Dr Margaret Hughes PhD[46] ,Dr David K Jackson [1] ,Dr Ben Jackson PhD [19] ,Dr Keith James [1] ,Natasha Johnson BSc (Hons)[21] ,Ian Johnston [1], Jon-Paul Keatley [1], Dr Moritz Kraemer [24], Dr Angie Lackenby [7], Dr Mara Lawniczak [1] , Dr David Lee [7], Rich Livett [1], Stephanie Lo [1], Daniel Mair BSc (Hons) [21], Joshua Maksimovic FD sport science [36], Nikos Manesis [7] ,Dr Robin Manley Ph.D [49], Dr Carmen Manso [7] ,Dr Angela Marchbank BSc [34] ,Dr Inigo Martincorena [1] ,Dr Tamyo Mbisa [7], Kathryn McCluggage MSC [36],Dr JT McCrone PhD [19], Shahjahan Miah [7] , Michelle L Michelsen BSc [49] , Dr Mari Morgan PhD [33], Dr Gaia Nebbia PhD, FRCPath [78],Charlotte Nelson MSc [46] ,Jenna Nichols BSc (Hons) [21] ,Dr Paola Niola PhD [41] ,Dr Kyriaki Nomikou PhD[21] ,Steve Palmer [1] , Dr. Naomi Park [1], Dr Yasmin A Parr PhD[21] ,Dr Paul J Parsons PhD [38] , Vineet Patel [7] ,Dr. Minal Patel [1] ,Clare Pearson MSc [2, 1] ,Dr Steven Platt [7] ,Christoph Puethe [1], Dr. Mike Quail [1],Dr JaynaRaghwani [24] , Dr Lucille Rainbow PhD [46] ,Shavanthi Rajatileka [1] ,Dr Mary Ramsay [7] ,Dr Paola C Resende Silva PhD [41, 42], Steven Rudder 51, Dr Chris Ruis [3] ,Dr Christine M Sambles PhD [49] ,Dr Fei Sang PhD [54] ,Dr Ulf Schaefer[7] ,Dr Emily Scher PhD [19] ,Dr. Carol Scott [1] ,Lesley Shirley [1], Adrian W Signell BSc [76] ,John Sillitoe [1] ,Christen Smith [1] ,Dr Katherine L Smollett PhD [21] ,Karla Spellman FD [36] ,Thomas D Stanton BSc [19] ,Dr David J Studholme PhD [49] ,Ms Grace Taylor-Joyce BSc [71] ,Dr Ana P Tedim PhD [51] ,Dr Thomas Thompson PhD[6] ,Dr Nicholas M Thomson PhD [51] ,Scott Thurston[1] ,Lily Tong PhD [21] ,Gerry Tonkin-Hill [1], Rachel M Tucker MSc [38] , Dr Edith E Vamos PhD [4],Dr Tetyana Vasylyeva[24] , Joanna Warwick-Dugdale BSc [49] , Danni Weldon [1] ,Dr Mark Whitehead PhD [46] ,Dr David Williams [7],Dr Kathleen A Williamson PhD[19],Harry D Wilson BSc [76],Trudy Workman HNC [34] ,Dr Muhammad Yasir PhD [51], Dr Xiaoyu Yu PhD [19], and Dr Alex Zarebski [24].

Samples and logistics:

Dr Evelien M Adriaenssens PhD [51], Dr Shazaad S Y Ahmad MSc [2, 47] , Adela Alcolea-Medina MPharm [59, 77] ,Dr John Allan PhD[60], Dr Patawee Asamaphan PhD[21], Laura Atkinson MSc [40], Paul Baker MD [63], Professor Jonathan Ball PhD [55], Dr Edward Barton MD[64], Dr. Mathew A Beale[1], Dr. Charlotte Beaver[1] , Dr Andrew Beggs PhD[16], Dr Andrew Bell PhD[51], Duncan J Berger [1], Dr Louise Berry. [56], Claire M Bewshea MSc [49], Kelly Bicknell [70], Paul Bird [58], Dr Chloe Bishop [7] , Dr Tim Boswell [56], Cassie Breen BSc[48], Dr Sarah K Buddenborg[1], Dr Shirelle Burton-Fanning MD [66,] Dr Vicki Chalker [7], Dr Joseph G Chappell PhD [55], Themoula Charalampous MSc [78, 94], Claire Cormie[3], Dr Nick Cortes PhD[29, 25], Dr Lindsay J Coupland PhD [52], Angela Cowell

MSc[48] , Dr Rose K Davidson PhD [53] , Joana Dias MSc[3] , Dr Maria Diaz PhD[51] , Thomas Dibling[1], Matthew J Dorman[1], Dr Nichola Duckworth[57], Scott Elliott[70], Sarah Essex[63], Karlie Fallon [58] , Theresa Feltwell [8] , Dr Vicki M Fleming PhD [56], Sally Forrest BSc [3], Luke Foulser[1], Maria V Garcia-Casado[1], Dr Artemis Gavriil PhD [41], Dr Ryan P George PhD[47], Laura Gifford MSc [33], Harmeet K Gill PhD[3], Jane Greenaway MSc[65], Luke Griffith Bsc[53], Ana Victoria Gutierrez[51], Dr Antony D Hale MBBS[85], Dr Tanzina Haque FRCPath, PhD[91], Katherine L Harper MBiol[85], Dr Ian Harrison [7] , Dr Judith Heaney PhD[89], Thomas Helmer [58], Ellen E Higginson PhD [3] , Richard Hopes [2], Dr Hannah C Howson-Wells PhD [56], Dr Adam D Hunter [1], Robert Impey [70], Dr Dianne Irish-Tavares FRCPath [91], David A Jackson[1] , Kathryn A Jackson MSc [46] , Dr Amelia Joseph [56], Leanne Kane [1], Sally Kay [1], Leanne M Kermack MSc [3], Manjinder Khakh [56], Dr Stephen P Kidd PhD[29, 25,31], , Dr Anastasia Kolyva PhD [51], Jack CD Lee BSc [40], Laura Letchford [1] , Nick Levene MSc[79], Dr LisaJ Levett PhD [89], Dr Michelle M Lister PhD [56], Allyson Lloyd [70] , Dr Joshua Loh PhD[60] , Dr Louissa R Macfarlane-Smith PhD[85], Dr Nicholas W Machin MSc [2 , 47], Mailis Maes M.phil[3], Dr Samantha McGuigan [1], Liz McMinn [1], Dr Lamia Mestek-Boukhibar D.Phil [41], Dr Zoltan Molnar PhD [6], Lynn Monaghan [79], Dr Catrin Moore [27], Plamena Naydenova BSc [3], Alexandra S Neaverson [1], Dr. Rachel Nelson PhD [1], Marc O Niebel MSc[21] , Elaine O'Toole BSc [48] , Debra Padgett BSc [64], Gaurang Patel [1] , Dr Brendan AI Payne MD [66], Liam Prestwood [1], Dr Veena Raviprakash MD[67], Nicola Reynolds PhD[86] Dr Alex Richter PhD [16], Dr Esther Robinson PhD[95], Dr Hazel A Rogers[1], Dr Aileen Rowan PhD [96], Garren Scott BSc [64], Dr Divya Shah PhD[40], Nicola Sheriff BSc [67], Dr Graciela Sluga MD - MSc[92] , Emily Souster[1], Dr. Michael Spencer-Chapman[1], Sushmita Sridhar BSc[1, 3], Tracey Swingler [53], Dr Julian Tang[58], Professor Graham P Taylor DSc[96], Dr Theocharis Tsoleridis PhD[55], Dr Lance Turtle PhD MRCP[46], Dr Sarah Walsh [57], Dr Michelle Wantoch PhD [86], Joanne Watts BSc[48] , Dr Sheila Waugh MD[66], Sam Weeks[41], Dr Rebecca Williams BMBS [31], Dr Iona Willingham[56], Dr Emma L Wise PhD [25, 29, 31], Victoria Wright BSc [54], Dr Sarah Wyllie [70] , and Jamie Young BSc [3].

Software and analysis tools

Amy Gaskin MSc[33], Dr Will Rowe PhD [15], and Dr Igor Siveroni PhD[96].

Visualisation:

Dr Robert Johnson PhD [96].

1 Wellcome Sanger Institute, **2** Public Health England, **3** University of Cambridge, **4** Health Data Research UK, Cambridge, **5** Public Health Agency, Northern Ireland ,**6** Queen's University Belfast **7** Public Health England Colindale, **8** Department of Medicine, University of Cambridge, **9** University of Oxford, **10** Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust; Cambridge, UK, **11** Division of Virology, Department of Pathology, University of Cambridge, **12** The Francis Crick Institute, **13** Cambridge Institute for Therapeutic Immunology and Infectious Disease, Department of Medicine, **14** Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK, **15** Institute of Microbiology and Infection, University of Birmingham, **16** University of Birmingham, **17** Queen Elizabeth Hospital, **18** Heartlands Hospital, **19** University of Edinburgh, **20** NHS Lothian, **21** MRC-University of Glasgow Centre for Virus Research, **22** Institute of Biodiversity, Animal Health & Comparative Medicine, University of Glasgow, **23** West of Scotland Specialist Virology Centre, **24** Dept Zoology, University of Oxford, **25** University of Surrey, **26** Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, **27** Big Data Institute, Nuffield Department of Medicine, University of Oxford, **28** Oxford University Hospitals NHS Foundation Trust, **29** Basingstoke Hospital, **30** Centre for Genomic Pathogen Surveillance, University of Oxford, **31** Hampshire Hospitals NHS Foundation Trust, **32** University of Southampton, **33** Public Health Wales NHS Trust, **34** Cardiff University, **35** Betsi Cadwaladr University Health Board, **36** Cardiff and Vale University Health Board, **37** Swansea University, **38** University of Sheffield, **39** Sheffield Teaching Hospitals, **40** Great Ormond Street NHS Foundation Trust, **41** University College London, **42** Oswaldo Cruz Institute, Rio de Janeiro **43** North

West London Pathology, **44** Imperial College Healthcare NHS Trust, **45** NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, **46** University of Liverpool, **47** Manchester University NHS Foundation Trust, **48** Liverpool Clinical Laboratories, **49** University of Exeter, **50** Royal Devon and Exeter NHS Foundation Trust, **51** Quadram Institute Bioscience, University of East Anglia, **52** Norfolk and Norwich University Hospital, **53** University of East Anglia, **54** Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, **55** Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, **56** Clinical Microbiology Department, Queens Medical Centre, **57** PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, **58** Clinical Microbiology, University Hospitals of Leicester NHS Trust, **59** Viapath, **60** Hub for Biotechnology in the Built Environment, Northumbria University, **61** NU-OMICS Northumbria University, **62** Northumbria University, **63** South Tees Hospitals NHS Foundation Trust, **64** North Cumbria Integrated Care NHS Foundation Trust, **65** North Tees and Hartlepool NHS Foundation Trust, **66** Newcastle Hospitals NHS Foundation Trust, **67** County Durham and Darlington NHS Foundation Trust, **68** Centre for Enzyme Innovation, University of Portsmouth, **69** School of Biological Sciences, University of Portsmouth, **70** Portsmouth Hospitals NHS Trust, **71** University of Warwick, **72** University Hospitals Coventry and Warwickshire, **73** Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCW NHS Trust, **74** Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, **75** Centre for Clinical Infection & Diagnostics Research, St. Thomas' Hospital and Kings College London, **76** Department of Infectious Diseases, King's College London, **77** Guy's and St. Thomas' Hospitals NHS Foundation Trust, **78** Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, **79** Princess Alexandra Hospital Microbiology Dept. , **80** Cambridge University Hospitals NHS Foundation Trust, **81** East Kent Hospitals University NHS Foundation Trust, **82** University of Kent, **83** Gloucestershire Hospitals NHS Foundation Trust, **84** Department of Microbiology, Kettering General Hospital, **85** National Infection Service, PHE and Leeds Teaching Hospitals Trust, **86** Cambridge Stem Cell Institute, University of Cambridge, **87** Public Health Scotland, 88 Belfast Health & Social Care Trust, **89** Health Services Laboratories, **90** Barking, Havering and Redbridge University Hospitals NHS Trust, **91** Royal Free NHS Trust, **92** Maidstone and Tunbridge Wells NHS Trust, **93** University of Brighton, **94** Kings College London, **95** PHE Heartlands, **96** Imperial College London, **97** Department of Infection Biology, London School of Hygiene and Tropical Medicine.

Appendix

Summary of major results from each study

A summary of each major study is presented below, grouped alphabetically by country.

South East Netherlands [37]

A study [37] of staff at LTCFs was done in March-April 2020 in south east Netherlands to understand their potential role in the spread of COVID-19 amongst residents in LTCFs. Staff members and a number of residents with mild respiratory symptoms were targeted for testing using rRT-PCR. The policy at the time was for staff with mild respiratory symptoms to wear a mask and still attend for work. At this point in time symptoms of COVID-19 and asymptomatic spread of SARS-CoV-2 were poorly understood.

A high prevalence of positive tests was found amongst staff at LTCFs, with 133 (21%) of those tested being positive; these were predominantly those providing care directly (nursing 74%, 1% physicians, 20% other health care workers) but also staff without patient contact (3%). Whilst the original intention was to test staff with mild respiratory symptoms, many staff with moderate symptoms were identified. Twenty-two samples from staff and residents at two LTCFs were selected for sequencing using an amplicon method, similar to ARTIC. Phylogenetic analysis indicated that the samples clustered by LTCF, with the exception of a single sample which represented a second introduction in one LTCF. Data is not in public archives and only available on request from the authors.

South Holland, Netherlands [33]

A study [33] was done to investigate an outbreak at a single LTCF in South Holland in the Netherlands from April to June 2020. A resident tested positive for SARS-CoV-2 after discharge from hospital, seeding an outbreak at the LTCF. This was confirmed by

sequencing 60 samples from staff and residents at the LTCF and from the epidemiologically linked hospital ward where the LTCF resident had been an inpatient. All genomes clustered together, with some forming two sub-clusters, confirming the relatedness of the outbreaks. Sequencing helped to understand transmission patterns which would not have been observed using traditional epidemiological techniques. Of the residents tested for SARS-CoV-2 using rRT-PCR, 113 (62.4%) were positive; only four declined to be tested. Residents who tested positive were more likely to be older and have cognitive impairment compared with those that tested negative.

Of the staff tested, 73 (20.8%) were positive; 34 (9%) declined to be tested. A large percentage of staff (65%) reported working while symptomatic. There was no difference in Ct values between symptomatic, pre-symptomatic and asymptomatic individuals, where pre-symptomatic is defined as individuals who were asymptomatic at the point of testing, but later developed symptoms. The sequence data is not available in the public archives and only available on request from the authors.

East of England study, UK [26]

A large-scale surveillance study of SARS-CoV-2 positive cases was done in the East of England, UK, between February and May 2020 [26], with genomic analysis of samples from 292 LTCFs (residential and nursing homes). This is considerably larger than previous studies and linked epidemiological data with genomic data. Included in the study were 7,406 samples that tested positive for SARS-CoV-2 from 6,600 patients. Of these 1,167 / 6,413 (18.2%) of the study population were residents in 337 LTCFs. From these 193 / 337 (57.3%) were residential homes and 144 / 337 (42.7%) were nursing homes, mainly located in five counties in the East of England (Cambridgeshire, Bedfordshire, Essex, Hertfordshire and Suffolk). This represents around half of the care homes in the region that had reported suspected or confirmed COVID-19 outbreaks to PHE at the time. Diagnostic samples were

tested at the Public Health England (PHE) Clinical Microbiology and Public Health Laboratory (CMPHL) at Cambridge University Hospitals NHS Foundation Trust. Samples were sequenced in the Division of Virology, Department of Pathology, University of Cambridge, as part of the COVID-19 Genomics Consortium UK (COG-UK).

The investigators found 409 distinct viral clusters in 292 LTCFs, corresponding to approximately half of all the LTCFs in the region that had reported outbreaks. Multiple potential transmissions between residents and staff were identified using genomic data. Multiple clusters per care home suggested that independent introductions were common and that within-care home transmission occurred frequently. The median number of cases per care home was two (range 1 - 22), with ten (3%) LTCFs with the highest numbers of cases accounting for 22% of all cases. There was evidence of large-scale outbreaks of identical or near identical (≤ 1SNP difference) lineages in care homes with the largest numbers of genomes. A median of eight SNPs separated genomes within LTCFs, compared with a median of nine SNPs for a random selection of non- LTCFs samples, indicating that genomic diversity amongst positive samples from LTCFs was similar to that from non-LTCFs samples. There were two LTCFs, located within 1km of each other, that had probable inter- LTCF transmission, with links to the same paramedics and shared carers. The genomes present in each had zero SNP differences, with only 2 days between the times that samples were taken. By combining epidemiological and genomic data it was possible to confirm a high probability of transmission. Admission and patient movement data were highlighted as a priority for investigation in relation to transmission.

The proportion of LTCF residents testing positive increased as general transmission decreased during lockdown, although it should be noted that sample collection strategies changed during the study period. Cases in LTCFs appeared more resistant to non-pharmaceutical intervention (NPI) measures. The study investigators also examined links between LTCFs cases and hospital admissions (Hamilton, personal communication). During

the study period 470 / 694 (67.7%) of LTCF residents had at least one hospital admission, and 398 / 694 (57.3%) were admitted to hospital with COVID-19 infection. 40 / 694 (5.8%) cases were categorised as suspected hospital-acquired COVID-19 infections. Furthermore, 230 / 694 (33.1%) of individuals were discharged from hospital within seven days of their first positive test, and could therefore have been infectious at the time of hospital discharge. Limiting the spread of SARS-CoV-2 between hospitals, healthcare workers and LTCF residents should be an urgent infection control and public health priority.

During the study period, no new viral lineages from outside the UK were observed in the entire dataset; this included genomes from LTCF and non-LTCF samples, suggesting travel restrictions had been successful in minimising new importations. The genome sequencing was based on the ARTIC protocol (https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w) utilising Nanopore and Illumina sequencing platforms as part of the COG-UK consortium. Overall, viral sequence data was not available for 40% of samples from LTCF residents testing positive; this was due to a combination of missing samples, mismatches between metadata and sequences, genomes not passing quality control, or sequencing being unavailable at the time of analysis. This highlights the practical difficulties in undertaking genomic surveillance when there are large numbers of samples. Despite availability of: all the consensus genomes in the database; the Global Initiative on Sharing All Influenza Data (GISAID) [31]; and al the raw data in the European Nucleotide Archive (ENA), the links between LTCF samples used for the analysis were not available (for reasons of patient confidentiality), although they may be requested from the authors.

Norfolk, UK [27]

A large-scale surveillance study [27] was undertaken in Norfolk, UK from March to August 2020 as part of the COG-UK consortium. A total of 42% (n=1,035) of all samples from SARS-CoV-2 positive cases from the hospital testing system within the region (covering

hospitals, LTCF, health care workers) were sequenced. An outbreak in one LTCF was investigated using data from genome sequencing of samples that had been prospectively collected. It was noted that the genomes from this LTCF were identical to each and formed a distinct sub lineage that included genomes from additional cases clustered in small geographical areas around the LTCF within a short period of time (between 8 April and 1 June 2020). As most of these additional cases were in older people (>85 years old), it was suspected that they represented LTCF clusters. It was later confirmed that the sub lineage was circulating in six LTCFs.

There were 90 cases caused by this sub lineage, of which 64 were known to be LTCF residents, nine were healthcare workers, and three were family members of healthcare workers. The majority of the LTCF infections were community-acquired. Twelve of these cases were admitted to a hospital (two were admitted twice to three different hospital trusts). Six had a community-acquired infection, testing positive within seven days of admission, three were inconclusive due to missing data, one had a probable hospital-acquired infection and tested positive within seven days of discharge, and two had a definite hospital-acquired infection (https://www.gov.uk/government/publications/wuhan-novel-coronavirus-infection-prevention-and-control/epidemiological-definitions-of-outbreaks-and-clusters-in-particular-settings). All LTCF residents with hospital-acquired infections were tested prior to discharge, suggesting that the adult social care IPC Department of Health and Social Care measures announced in May (which required this) were being followed. In the time period covered by the study, while patients required a test before they could be discharged to a LTCF, a positive test did not preclude them from returning to the LTCF just that adequate IPC measures needed to be taken. Given that some of this cohort of patients tested positive for community-acquired infections in May, a number of weeks after the adult social care IPC Department of Health and Social Care measures were announced, suggests that these measures may not have been sufficient.

This multi- LTCF outbreak was only identified using genomic epidemiology and the links between LTCFs were unknown prior to this investigation. The sub lineage involved was not detected in cases identified as part of community testing. This study indicated inter- LTCF transmission was likely. The initial seeding events were unknown. A limitation of the study is that the collection date of a sequenced sample may not be the first positive test for that case. All data analysed as part of this study is detailed in the preprint, with accession numbers for each sample linked to accession numbers for the public archives, which enables reanalysis.

London four study, UK [20]

An investigation of an outbreak in four LTCFs in London was done in April 2020 [20]. Residents were tested for SARS-CoV-2 by rRT-PCR at two time points one week apart; residents who were initially positive were not re-tested. Of all those residents tested 126 (40%) were positive; of these 54 (42.9%) were asymptomatic. There was a high COVID-19 mortality rate (26%) amongst residents during the study period. In terms of ethnicity, 18.5% of the residents were Black, Asian and minority ethnic (BAME) and the mortality rate for these residents was similar to that of white residents.

A subset (n=70, 11%) of asymptomatic staff at three of the four LTCFs were offered SARS-COV-2 testing. Only three (4%) out of 70 tested positive, but it was noted that staff absence rates due to sickness or self-isolation were more than twice the normal rate during the study period. Furthermore, three different diagnostic platforms were used for testing, including one with a lower sensitivity than the others. These changes in methodology may have affected the reliability of the study results.

In terms of sequencing, samples from one staff member and 17 (19 reported in the manuscript) residents were sequenced. However, samples that were sequenced were not

representative of all the LTCFs; there was only one sample sequenced from one LTCF and only two samples sequenced from a second LTCF. The bioinformatics methods used (assembling amplicons) is generally regarded as poor practice and likely to result in errors in the sequences generated. Furthermore, no sequencing data were publicly available, precluding reanalysis of the primary data.

The genomic data identified a cluster with one staff member and two residents at a single care home. Most of the samples clustered by care home, and visual inspection of the phylogenetic tree presented indicates there were two clusters in each of two care homes, although this was not entirely clear from the data presented. The data were compared to a random selection of UK samples to provide background context and help to show separation between clusters.

London six study, UK [21]

A study of six London (UK) LTCFs experiencing COVID-19 outbreaks was done over the Easter weekend (10-12 April, 2020) [21,38]. The LTCSs in this region had outbreaks of COVID-19 early on in the UK epidemic, before there was full recognition of the extent of community transmission and the frequency of asymptomatic transmission. The study was of 518 individuals and included both staff and residents; 105 residents and 53 staff tested positive for SARS-CoV-2 using rRT-PCR. Virus was cultured by Public Health England Colindale to ascertain infectivity and patients were tested serologically for evidence of previous infection [35].

Of those who tested positive for SARS-CoV-2 there was a high percentage of asymptomatic cases amongst staff (49%, n=26) and residents (44%, n=46), indicating that symptom screening has low sensitivity. Detection of outbreaks was often delayed if based on symptoms as, by then there were already high rates of asymptomatic infection in both staff

and residents. The rRT-PCR cycle threshold values (Ct), which indicate SARS-CoV-2 viral load, were similar across different age groups, and between symptomatic and asymptomatic cases. Infectivity in culture was also similar across different age groups, and between symptomatic and asymptomatic cases. A high percentage of cases with symptoms tested negative by rRT-PCR (15%, n=24 residents; 9%, n=19 staff). This may indicate that: sampling was inadequate; viral loads were too low to be detected (early/late infection); or that diagnostic screening by rRT-PCR with a single target gene may underestimate infection. Residents who were symptomatic and tested positive by rRT-PCR had a higher mortality rate than those who were symptomatic and tested negative by rRT-PCR (36% versus 4%).

Genome sequencing of 99 out of 158 cases that tested positive (62%) revealed two distinct lineages [39] predominantly B.1 and B.2.1, two of the most common UK lineages. All six LTCFs had both lineages, and genomes from both staff and residents were interspersed throughout the phylogenetic tree, likely due to the low genetic diversity of SARS-CoV-2 genomes in April, 2020. To provide genomic context we examined the publicly available virus genomes from the COVID-19 Genomics Consortium UK that had been collected in the week before and the week after the Easter weekend (5 - 19 April 2020) in the Greater London region. At this time, diagnostic testing was directed towards people with symptoms, so may not have been representative of community spread, but did provide an indication of the virus diversity within this small geographic region. A total of 44 lineages were observed from 617 genomes, including the London Six genomes. However, the London Six publications did not provide sample accession numbers so it was not possible to identify them within the public archives. The most common lineages in Greater London at the time were B.1.1.1 (n=298, 48%) followed by B.1 (n=90, 14%), B.2.1 (n=78, 13%) and B.1.5 (n=21, 3%).

The study found that there were up to nine separate introductions into a single LTCF. Reanalysis for this review indicates that there was more sequence diversity than expected

for samples from the Easter time period, and the number of introductions into a single LTCF was likely to be six rather than nine. This over-estimation of introductions and sequencing diversity was caused by some poor-quality sequence data, including missing data, leading to bioinformatics artefacts. Had there been a high level of introductions into a single LTCF we would expect to observe more lineages, particularly the most common lineage for the region. The raw sequencing data and genomes are available in the public archives, however the specific samples used for this study were not detailed in the papers to maintain patient confidentiality. This limits public reanalysis.

Boston, USA [25]

A large community surveillance study was conducted between January and May 2020 in Boston, USA [25]. In this study 850 SARS-CoV-2 positive samples underwent direct sequencing using Illumina, with reference-guided assembly. Over 80 introductions were estimated to have occurred in the region over the study period.

A sub-study conducted in April 2020 analysed an outbreak in a single LTCF, where a planned relocation of residents led to universal screening of residents and staff for SARS-CoV-2. Out of those tested 82 (85%) residents and 36 (37%) staff tested positive. A total of 83 (67%) genomes were sequenced. From these 75 (90%) genomes formed a single closely-related cluster; 59 were identical (no SNP differences) and shared a distinct mutation (G3892T) with unknown significance. Genome sequencing indicated a recent introduction from a single source. Estimates for the most recent common ancestor allowed the authors to estimate that the time from introduction to widespread positive testing in residents was 2 - 3 weeks. Two additional introductions (three genomes each) were also observed but did not disseminate widely. The three introductions highlight the risk of introduction into a high-risk setting, despite strict infection control measures which had been in place from two weeks prior to the estimated introduction date. By tracking the mutation distinct to this outbreak as

part of continued regional surveillance and sequencing indicated that there was little onward spread from this initial superspreading event. The raw read sequence data and the assembled genomes are deposited in the public archives (NCBI), allowing for reanalysis.

California, USA [29]

A four-week prospective surveillance study was done on 192 patients with COVID-19 in a hospital in Los Angeles, California, USA between March and April 2020 [29]. Genome sequencing found that 85% of genomes were European lineages and 15% were Asian, indicating multiple sources of introduction. Out of all the samples, 113 (69%) yielded genomes of sufficient quality for use in phylogenetic analysis (>50% reconstructed consensus genome). The percentage of the genome that could be reconstructed was closely correlated with the number of viral copies in the primary sample.

From phylogenetic analysis of the sequenced SARS-CoV-2 genomes, a cluster of ten patients was identified: five of these were residents from a single LTCF, while the other five were associated with a LTCF one block away (three staff members, a family member of a resident and one resident). Another related case was identified in a person living near one of the LTCFs. Genome sequencing was used to establish connections between these cases; the genomes were identical (or near identical) to each other and belonged to lineage B.1. In total the study identified three large clusters, only one of which included genomes from a LTCF. This study demonstrates the effectiveness of prospective surveillance in detecting and linking outbreaks in LTCFs, and thus enhancing contact tracing efforts. The data are deposited in GISAID, with samples clearly described allowing for reanalysis.

Colorado, USA [16]

A prospective surveillance study of staff at five LTCFs was done over a six-week period in Colorado, USA [16]. This involved consecutive testing for SARS-CoV-2 in staff at five LTCFs

to investigate the prevalence of asymptomatic and pre-symptomatic positive tests. Staff voluntarily enrolled and were swabbed weekly throughout the study period, including if they developed symptoms. Staff with and without direct contact with residents were included in the study. A total of 70 staff members tested positive and rates of infection varied between LTCFs. The median number of consecutive weekly positive tests was 2 (range 1 to 5), indicating a detection window in the nasopharynx of most people of at least 8 days. Some individuals tested positive for five consecutive weeks, and some tested positive intermittently. The levels of viral RNA tended to decline over the duration of infection and corresponded to low levels of infectious virus in culture.

A total of 48 genomes from positive samples were sequenced, ten of which came from five staff members collected over two consecutive weeks. The ARTIC amplicon protocol was used, with Illumina sequencing; gaps in the consensus sequences were filled with bases from the reference genome, so the results should be treated with caution. Of those sequenced, 36 genomes from one LTCF clustered together, and a further five clustered with another LTCF. Transmission within the workplace was likely, but community transmission could not be ruled out. Of the five staff members with two sequenced genomes each, three had genomes that differed in SNPs between the two consecutive samples; this high rate of within-host mutation is likely to be due to a bioinformatics error associated with filling gaps with bases from the reference genome. The sequence data are not publicly available and could not be reanalysed.

Minnesota, USA [32]

A prospective surveillance study was done in two LTCFs experiencing COVID-19 outbreaks between April and June 2020 in Minnesota, USA [32]. Residents (n=261) and staff (n=480) were offered SARS-CoV-2 testing up to six times and, once they tested positive, were not re-tested. Participation rates in testing varied by LTCF, with 17% of residents in one refusing

to be tested initially. Of those that were tested, 165 (64%) residents tested positive, 33 of these (20%) were hospitalised and 52 (31%) died. Residents testing positive were isolated in a COVID-19 specific unit, but this had no impact on overall transmission as indicated by the continued identification of positive cases throughout the study. This study demonstrated the utility of serial (repeated) sampling of the same individuals for detecting new cases as they occurred, with the detection of new cases rapidly diminishing throughout the study.

Severe challenges were encountered with staff testing. Staff were reluctant to participate (71%), and when they did, did so only once. Overall, 114 (33%) staff tested were positive, of which 58 (51%) were symptomatic and working on the day of testing. Of those staff testing positive, 41 (12%) were not involved directly in care provision. There were delays of up to 12 days in obtaining test results, with staff incurring financial losses if they self-isolated without a positive test. Four staff members were hospitalised and two staff members died.

In this study genomes from 105 samples were sequenced using the ARTIC protocol. Genomes were clustered into two groups separated by LTCF (i.e all genomes in one cluster were from residents and staff of one LTCF while all genomes in the other cluster were from residents and staff of the other LTCF); this indicates within-home transmission and no evidence for transmission between LTCFs. Only 37% of positive samples were available for sequencing; samples from early in the outbreaks were missing. However, this was sufficient to be reasonably confident of the underlying clusters and dynamics observed. In one LTCF there appeared to be a second potential introduction event from the community, rooted earlier in the tree. However, as there were few specimens from early stages of the outbreak for sequencing, the full genetic evolutionary history cannot be elucidated further. Sequence data from this study are available on GISAID, but there are no sample identifiers or accession numbers in the manuscript to enable linkage or re-analysis of the data.

Washington, USA [17]

This study in Washington state, USA, [17] was the first to report the use of genome sequencing to investigate a large COVID-19 outbreak in a LTCF. Following positive results from SARS-CoV-2 testing of one staff member and one resident, samples from residents were tested by Public Health–Seattle and King County (PHSKC) and the Centers for Disease Control and Prevention (CDC) on two occasions one week apart. Not all residents were tested. Real-time reverse transcription polymerase chain reaction (rRT-PCR) was used for identification and a subsample of positive samples were selected for culture and sequencing. No new residents were admitted to the LTCF after the first resident tested positive. Enhanced infection prevention and control (IPC) measures focused on symptomatic residents and staff were implemented after the first resident tested positive. However, testing 3 days after implementation showed widespread transmission had already occurred. There was a high percentage of positive cases amongst residents (64%, n=57), most of whom were asymptomatic or pre-symptomatic at the time of testing (82%, n=48). Fifteen residents died (26%) and a further 11 were hospitalised. No serological testing was done. The viral load (based on the rRT-PCR cycle threshold values [Ct]) and the percentage of cultures testing positive for virus were similar between symptomatic, pre-symptomatic and asymptomatic cases. Symptomatic staff (40%; n = 138) were advised to seek testing externally by their health care provider; of these 19% tested positive (n=26). Asymptomatic staff were not advised to be tested; this may have underestimated the infection rate. The role of staff in the introduction or transmission of SARS-CoV-2 was not fully explored or analysed in the study.

The doubling time was faster than in the surrounding community, but this may have been due to the identification of asymptomatic cases within the LTCS; only symptomatic individuals were tested in the community. The IPC measures focused on symptomatic cases, but the high prevalence of underlying conditions (cognitive impairment, chronic

coughs) amongst residents made identification of COVID-19 symptoms difficult, particularly in the early stages of infection.

Nanopore sequencing was done on samples from 34 residents that tested positive; for five of these residents samples were taken twice, one week apart, sequenced, and the viral genomes found to be identical in the second test in each case. This demonstrated the reproducibility of SARS-CoV-2 genome sequencing. Bioinformatics analysis showed that 79% (n=27) of positive samples mainly clustered into two groups, separated by a single SNP difference; a small number of outlier samples containing up to 4 SNP differences. This confirmed the relatedness of genomes found in the residents' samples. Identical sequences were given a unique identifier and, when these were related back to a map of the facility and the location of the residents' bedrooms, there was a very clear spatial signal; residents in adjacent bedrooms were more likely to have 100% identical consensus genomes than not. Phylogenetic analysis of publicly available genomes at the time showed that the genomes from the LTCS samples were very closely related to those found elsewhere in the locality (Washington, USA). Sequencing data used in the paper has been publicly deposited in two archives, with sufficient information in the paper to enable the genomic analysis to be fully reproduced.

Other studies

There is one study from Hungary [40] that sequenced a single LTCF resident's positive SARS-CoV-2 sample.

An unpublished study (personal communication Guthrie, Templeton and Holden) sequencing SARS-CoV-2 positive samples from staff and residents of LTCFs in Scotland as part of the COG-UK consortium. There was evidence for a connection between the genomes from staff

and residents' samples in some LTCFs. Outbreaks were heterogenous in size, duration and pattern (some explosive, some more drawn out some with long gaps between cases).

Another unpublished study (personal communication Bashton, Young, Nelson, Smith) done as part of the COG-UK consortium sequenced genomes from staff and residents testing positive at 64 LTCFs in the North Yorkshire, South Tees region. Of these LTCFs, 36 had multiple positive samples enabling genome sequencing and cluster analysis. Sequence data analysis using Civet (https://github.com/artic-network/civet) detailed six outbreak clusters. One of these clusters involved three LTCFs and associated staff from their local National Health Service (NHS) trust. Another involved two LTCFs and an associated staff member. This demonstrated not only transmission between residents within LTCFs, but more complex transmission chains between LTCFs and local hospitals.

Definitions

We use the term 'long term care facility' which encompasses terms used to describe similar facilities in different countries such as: 'skilled nursing facility', 'care home', 'nursing home', 'elderly care home', and 'residential home'.

For the purposes of this review we use the ECDC definition of nosocomial infection (https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions).