THE LANCET
Digital Health

Supplementary appendix

Supplement to: Thygesen JH, Tomlinson C, Hollings S, et al. COVID-19 trajectories among 57 million adults in England: a cohort study using electronic health records. *Lancet Digit Health* 2022; published online June 8. https://doi.org/10.1016/S2589-7500(22)00091-7.

Supplementary Material - COVID-19 trajectories among 57 million adults in England: a cohort study using electronic health records

Data sources and data quality

Primary care

Data in primary care have been sourced from the General Practice Extraction Service (GPES) Data for pandemic planning and research (GDPPR) system which contains SNOMED-CT concepts for patients registered with a primary care physician in the UK. The dataset contains approximately 96% of the English populations and 98% of all English general practices. Patients records were included in GDPPR when they had coded information matching any of the SNOMED-CT concepts in the Code Clusters applicable for COVID-19 planning during primary care consultations [1]. Around 34,000 unique SNOMED-CT concepts are included (>90% of all those currently extracted for a wide range of purposes by NHS Digital's GP Extraction Service), covering a broad range of diagnoses and procedures (from the start of each person's records) along with laboratory results, physical measurements, clinical referrals, and prescriptions. Primary care EHR have been shown to have a high degree of diagnostic accuracy in validation studies [2]

Hospitalisation

Hospital Episode Statistics (HES) contains administrative data from English hospitals in the National Health Service (NHS). HES captures a) inpatient episodes (including maternity), b) outpatient episodes, c) accidents and emergency attendance (A&E), d) critical care, and e) adult mental health. The primary purpose of HES is to facilitate hospital reimbursement which is actioned through a framework called "Payments by Results")[3].

HES captures the records of all patients, and their interactions, if they are funded by the NHS irrespective of if they are UK residents or if the care was delivered by an NHS provider. HES record-level data are structured into spells (admissions) which in turn are composed of one or more episodes (most admissions have a single episode)[4]. An episode can be defined as a period of continuous care from a single consultant/speciality and HES contains a row per episode per admission. A spell is terminated when a patient is discharged or dies. HES inpatient data are recorded using WHO ICD-10 and procedures using the Office of Population Censuses and Survey's (OPCS) version 4 clinical classification. A spell can have up to 20 primary and secondary diagnoses or procedures recorded. A primary diagnosis in HES is defined as the main condition treated (or investigated) during the episode of care or where no such definitive diagnosis exists, the main symptom or abnormal finding observed.

HES data are published annually but data are collected through a mechanism known as the Secondary Uses Service (SUS) which curates monthly data extracts from healthcare providers.

These monthly extracts are then subsequently used to populate the HES database. While variation is likely to exist between healthcare provides, coded data from hospitalisations (through Hospital Episode Statistics Admitted Patient Care and Secondary Uses Service have been shown to be robust: median diagnostic accuracy 80.3% (IQR: 63.3-94.1%) and median procedure accuracy of 84.2% (IQR: 68.7-88.7%)[5]. HES and SUS undergo robust data quality controls and validation rules which are further described along with the data processing pipeline elsewhere[6].

Mortality

In England (and Wales), when a patient dies, it is the statutory duty of the doctor who had attended in the last illness to issue the death certificate and the Office of National Statistics (ONS) centrally collects and curates all deaths. ONS mortality statistics are considered to be the gold standard for death ascertainment and are routinely used in EHR studies to ascertain deaths. During the pandemic however, there were a variety of changes to the processes in which deaths were certified and registered. For example, the time taken for deaths to be registered decreased while the numbers of conditions recorded on the death certificate were greater for deaths involving COVID-19 than those not involving COVID-19, suggesting higher rates of comorbidities in these deaths and good quality of the certification.[7]

Data Linkage

Individual data sources are linked by NHS Digital using the Master Person Service in combination with the Personal Demographics Service. A linkage score is calculated by cross-referencing information across different sources with the demographics in the Personal Demographics Service and signifies the overall associated match confidence. This score is not directly made available to researchers but NHS Digital's monthly reports for data quality maturity index indicate that 97-100% of records submitted to NHS Digital each month include information on NHS number and other key personal variables, providing confidence in the accuracy of the matching process [8].

Comparison with population estimates

The dataset comprises more than 96% coverage of the English population and represents the English population in terms of age, sex, ethnicity, and diabetes when compared with UK government official statistics for England, includes the full distribution of general practices according to geographical location and size. The datasets and their underlying characteristics as described in detail elsewhere[9].

None of the datasets included patients that have explicitly opted out of their EHR being used for medical research. These are referred to as "Type 1 opt-outs" and as of Sept 1 2021 there were 3,264,327 national data opt-outs [10]. Lastly, this is a dynamic cohort where new patients can enter (at birth or new registration with a general practice) during the study period. For example, there are approximately 1 million migrations and student visitors in the UK yearly that are likely to register with a general practitioner but not be adequately captured in the ONS population estimates[11]. As a result, the total number of participants is a cumulative estimate over the study period and does not entirely align with national population estimates [12].

COVID-19 event phenotyping methodology

COVID-19 positive tests were defined as a positive result from national testing data (SGSS), encompassing tests from NHS hospitals for those with a clinical need and healthcare workers (known as ‘Pillar 1’) and swab testing from the wider population (known as ‘Pillar 2’). The timeliness and completeness of SGSS has been evaluated and shown to be high [13].

COVID-19 primary care diagnoses were identified from primary care (GDPPR) using SNOMED-CT terms. SNOMED-CT terms are recorded during primary care consultations by the general practitioner.

COVID-19 hospital admissions were defined as any hospital admission recorded in CHESS (COVID-19 specific hospitalisations data) or admissions with a COVID-19 ICD-10 codes in HES APC or SUS as primary or other listed cause of hospitalisation.

Provision of ventilatory support was ascertained from several sources:
a) patients with an ICU admission recorded in COVID-19 Hospitalisation in England Surveillance System (CHESS) or patients with an ICU admission (defined as an entry within HES Adult Critical Care) within the same admission as HES APC,

b) patients receiving NIV, identified using the OPCS4 code E85.2 (Non-invasive ventilation NEC) or E85.6 (Continuous positive airway pressure), from SUS or HES APC, or a positive number of days receiving basic respiratory support in HES Adult Critical Care, or ‘high flow nasal oxygen’ or ‘non invasive mechanical ventilation’ recorded in CHESS,

c) patients receiving Intermittent Mandatory Ventilation (IMV) were identified using OPCS-4 code E85.1 (Invasive ventilation) or X56 (Intubation of trachea), in SUS or HES APC, or a non-zero entry for days receiving advanced respiratory support from HES Adult Critical Care, or ‘invasive mechanical ventilation’ recorded in CHESS,

d) patients receiving Extracorporeal Membrane Oxygenation (ECMO) identified using OPCS4 X58.1 (Extracorporeal membrane oxygenation) in SUS or HES APC or ‘respiratory support ECMO’ recorded in CHESS.

Fatal COVID-19 events were identified from ONS Civil Registration of Deaths and secondary care (HES APC, SUS) and defined as:

a) a suspected or confirmed COVID-19 diagnosis ICD-10 term present in any position on the death certificate,

b) death within 28-days of the first recorded COVID-19 event (positive test, diagnosis or admission), irrespective of the cause of death recorded on the death certificate, or

c) a COVID-19 hospital admission with a discharge method (dismeth) or destination (disdest) denoting death, irrespective of cause and duration after the index event. See Supplementary Table 2 for the specific codes used.

Phenotype definitions were reviewed by clinicians, health data scientists and epidemiologists and validated by quantifying cross-EHR source concordance and checking consistency of findings with established risk factors from the literature, in keeping with the CALIBER approach[14].

Ethical and Regulatory Approvals

Data access approval was granted to the CVD-COVID-UK consortium (under project proposal CCU013 High-throughput electronic health record phenotyping approaches) through the NHS Digital online Data Access Request Service [15] (ref. DARS-NIC-381078-Y9C5K). For full detail see supplementary methods. The BHF Data Science Centre approvals and oversight board deemed that this project's work fell within the scope of the consortium’s ethical and regulatory approvals. Analyses were conducted by three approved researchers (JHT, CT, SD) via secure remote access to the TRE. Only summarised, aggregate results were exported, following manual review by the NHS Digital ‘safe outputs’ escrow service, to ensure no output placed in the public domain contains information that may be used to identify an individual [9]. The North East-Newcastle and North Tyneside 2 research ethics committee provided ethical approval for the CVD-COVID-UK research programme (REC No 20/NE/0161).

Collaborators

List of CVD-COVID-UK/COVID-IMPACT Consortium members and primary institutional affiliation as of 3rd of March 2022.
Authors of the paper are highlighted in yellow

Member Name	**Institution**
Abdel Douiri	King's College London
Abdelaali Hassaine	University of Oxford
Abdul Qadr Akinoso-Imran	Queen's University Belfast
Abraham Olvera-Barrios	University College London
Adejoke Oluyase	King's College London
Adnan Tufail	University College London
Ajay Shah	King's College London
Alan Carson	University of Edinburgh
Alasdair Warwick	University College London
Alastair Denniston	INSIGHT
Alastair Proudfoot	Barts Health NHS Trust
Alex Handy	University College London
Alexandru Dregan	King's College London
Alun Davies	Imperial College London
Alvina Lai	University College London
Amanj Kurdi	University of Strathclyde
Ami Banerjee	University College London
Amir Gavrieli	University of Cambridge
Ana Torralbo	University College London
Andrew Lambarth	University College London
Angela Henderson	University of Glasgow
Angela Wood	University of Cambridge
Anna Bone	King's College London
Anna Hansell	University of Leicester
Annemarie Docherty	University of Edinburgh
Anthony Khawaja	University College London
Antonella Delmestri	University of Oxford

Aoife McCarthy	University of Edinburgh
Arun Karthikeyan Suseeladevi	University of Bristol
Arun Pherwani	University Hospital of North Midlands
Ashkan Dashtban	University College London
Ashley Akbari	Swansea University
Badar Ahmed	NHS Digital
Baljean Dhillon	University of Edinburgh
Ben Bray	King's College London
Ben Goldacre	University of Oxford
Ben Humberstone	Office for National Statistics
Ben Lacey	University of Oxford
Bilal Mateen	Wellcome Trust
Brian Roberts	NHS Digital
Cameron Razieh	University of Leicester
Camille Harrison	Office for National Statistics
Carmen Petitjean	University of Cambridge
Carole Morris	NHS Scotland
Caroline Dale	University College London
Caroline Jackson	University of Edinburgh
Caroline Rogers	Healthcare Quality Improvement Partnership
Cathie Sudlow	Health Data Research UK / BHF Data Science Centre
Charles Wolfe	King's College London
Christian Schnier	University of Edinburgh
Christopher Tomlinson	University College London
Claire Lawson	University of Leicester
Claire Tochel	University of Edinburgh
Clare Gillies	University of Leicester
Clea du Toit	University of Glasgow
Colin Berry	University of Glasgow
Costas Kallis	Imperial College London
Craig Melville	University of Glasgow
Craig Smith	University of Manchester
Dan O’Connell	British Heart Foundation
Dani Prieto-Alhambra	University of Oxford

Daniel Harris	Swansea Bay University Health Board
Daniel Morales	University of Dundee
David Brind	University of Cambridge
David Cromwell	Royal College of Surgeons of England
David Hughes	University of Liverpool
David Jenkins	University of Manchester
David Moreno Martos	University of Dundee
David Selby	University College London
Deborah Kinnear	University of Glasgow
Deborah Lawler	University of Bristol
Deborah Lowe	NHS England
Deepti Gurdasani	Queen Mary University of London
Dexter Canoy	University of Oxford
Dominic Oliver	King's College London
Efosa Omigie	NHS Digital
Elena Nikiphorou	King's College London
Elena Raffetti	University of Cambridge
Elias Allara	University of Cambridge
Elizabeth A Ellins	Swansea University
Eloise Withnell	University College London
Elsie Horne	University of Bristol
Emanuele Di Angelantonio	University of Cambridge
Emma Copland	University of Oxford
Eva Morris	University of Oxford
Evaleen Malgapo	University College London
Evan Kontopantelis	University of Manchester
Ewan Birney	European Bioinformatics Institute
Fabian Falck	University of Cambridge
Fatemeh Torabi	Swansea University
Felix Greaves	NICE
Filip Sosenko	University of Glasgow
Flavien Hardy	University College London
Florian Falter	Royal Papworth Hospital NHS Foundation Trust
Francesco Zaccardi	University of Leicester

Frank Kee	Queen's University Belfast
Frederick Ho	University of Glasgow
Freya Allery	University College London
Gareth Davies	Swansea University
Gareth Williams	King's College London
Genevieve Cezard	University of Cambridge
George Nicholson	University of Oxford
George Tilston	University of Manchester
Gwenetta Curry	University of Edinburgh
Hannah Whittaker	Imperial College London
Haoting Zhang	University of Cambridge
Harry Hemingway	University College London
Harry Watson	King's College London
Harry Wilde	University of Warwick
Hoda Abbasizanjani	Swansea University
Honghan Wu	University College London
Howard Tang	University of Cambridge
Huan Wang	University of Dundee
Huayu Zhang	University of Edinburgh
Ify Mordi	University of Dundee
Irene Higginson	King's College London
Jake Kasan	NHS Digital
James Sheppard	University of Oxford
Jane Lyons	Swansea University
Javiera Leniz Martelli	King's College London
Jayati Das-Munshi	King's College London
Jen-Yu Amy Chang	University of Sheffield
Jennifer Beveridge	NICE
Jennifer Cooper	University of Bristol
Jennifer Quint	Imperial College London
Jennifer Rossdale	Royal United Hospitals Bath NHS Foundation Trust
Jessica Barrett	University of Cambridge
Jianhua Wu	University of Leeds
Jill Pell	University of Glasgow

Jinge Wu	University College London
Joanna Davies	King's College London
Johan Thygesen	University College London
Johannes Heyl	University College London
John Danesh	University of Cambridge
John Dennis	University of Exeter
John Macleod	University of Bristol
John Nolan	Health Data Research UK / BHF Data Science Centre
Johnny Downs	King's College London
Jon Boyle	Addenbrooke's Hospital
Jon Shelton	Cancer Research UK
Jonathan Sterne	University of Bristol
Joseph Firth	University of Manchester
Joshua Day	NHS Digital
Julia Hippisley-Cox	University of Oxford
Julia Townson	Cardiff University
Julian Halcox	Swansea University
Kamlesh Khunti	University of Leicester
Kate Cheema	British Heart Foundation
Katherine Brown	Great Ormond Street Hospital
Katherine Sleeman	King's College London
Katie Harron	University College London
Kazem Rahimi	University of Oxford
Ken Li	University College London
Kim Kavanagh	University of Strathclyde
Lamiece Hassan	University of Manchester
Lana Bojanić	University of Manchester
Lars Murdock	Cancer Research UK
Laura North	Swansea University
Laura Pasea	University College London
Livia Pierotti	University of Bristol
Lorna Fraser	University of York
Luanluan Sun	University of Cambridge
Luca Grieco	University College London

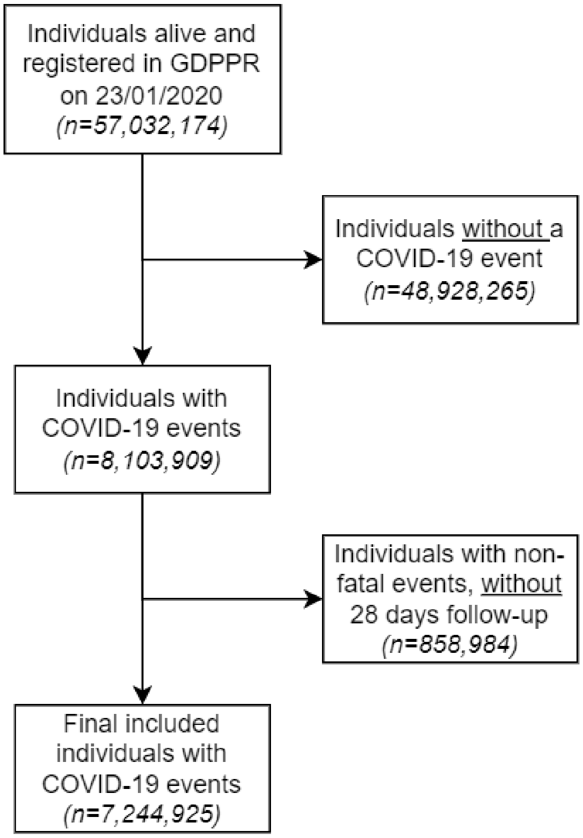
Lucy Wright	University of Oxford
Luisa Zuccolo	University of Bristol
Lynn Morrice	Health Data Research UK / BHF Data Science Centre
Mamas Mamas	Keele University
Maria Sudell	University of Liverpool
Marion Bennie	University of Strathclyde
Mark Ashworth	University of Oxford
Mark Barber	NHS Lanarkshire
Mark Green	University of Liverpool
Marta Pineda Moncusi	University of Oxford
Martha Elwenspoek	University of Bristol
Martina Slapkova	Cancer Research UK
Mary Joan Macleod	University of Aberdeen
Massimo Caputo	University of Bristol
Matt Sydes	University College London
Matthew Sperrin	University of Manchester
Maya Buch	University of Manchester
Mehrdad Mizani	University College London
Mevhibe Hocaoglu	King's College London
Michael Sweeting	University of Leicester
Michalis Katsoulis	University College London
Michelle Williams	University of Edinburgh
Miguel Bernabeu Llinares	University of Edinburgh
Mike Inouye	University of Cambridge
Milad Nazarzadeh Larzjan	University of Oxford
Mira Hidajat	University of Bristol
Mirek Skrypak	Healthcare Quality Improvement Partnership
Mohammad Mamouei	University of Oxford
Moritz Gerstung	European Bioinformatics Institute
Munir Pirmohamed	University of Liverpool
Myer Glickman	Office for National Statistics
Naomi Herz	British Heart Foundation
Naveed Sattar	University of Glasgow
Nazrul Islam	University of Oxford

Neil Davies	University of Bristol
Nick Hall	University of Oxford
Nida Ahmed	Barts Health NHS Trust
Nilesh Samani	British Heart Foundation
Norman Briffa	University of Sheffield
Olena Seminog	University of Oxford
Owen Pickrell	Swansea University
Paula Lorgelly	University College London
Pedro Machado	University College London
Pia Hardelid	University College London
Qiuju Li	London School of Hygiene & Tropical Medicine
Rachel Cripps	King's College London
Rachel Denholm	University of Bristol
Raph Goldacre	University of Oxford
Raymond Carragher	Queen's University Belfast
Rebecca Crallan	Cancer Research UK
Rebecca Milton	Cardiff University
Reecha Sofat	University of Liverpool
Renin Toms	University of Bristol
Richard Chin	University of Edinburgh
Richard Killick	King's College London
Richard Williams	University of Manchester
Riyaz Patel	University College London
Rocco Friebel	London School of Economics & Political Science
Rochelle Knight	University of Bristol
Rohan Takhar	University College London
Ronan Lyons	Swansea University
Rosie Hinchliffe	Cancer Research UK
Rouven Priedon	Health Data Research UK / BHF Data Science Centre
Rowena Griffiths	Swansea University
Roy Schwartz	University College London
Rupert Payne	University of Bristol
Russell Healey	NHS Digital
Rutendo Mapeta	University of Cambridge

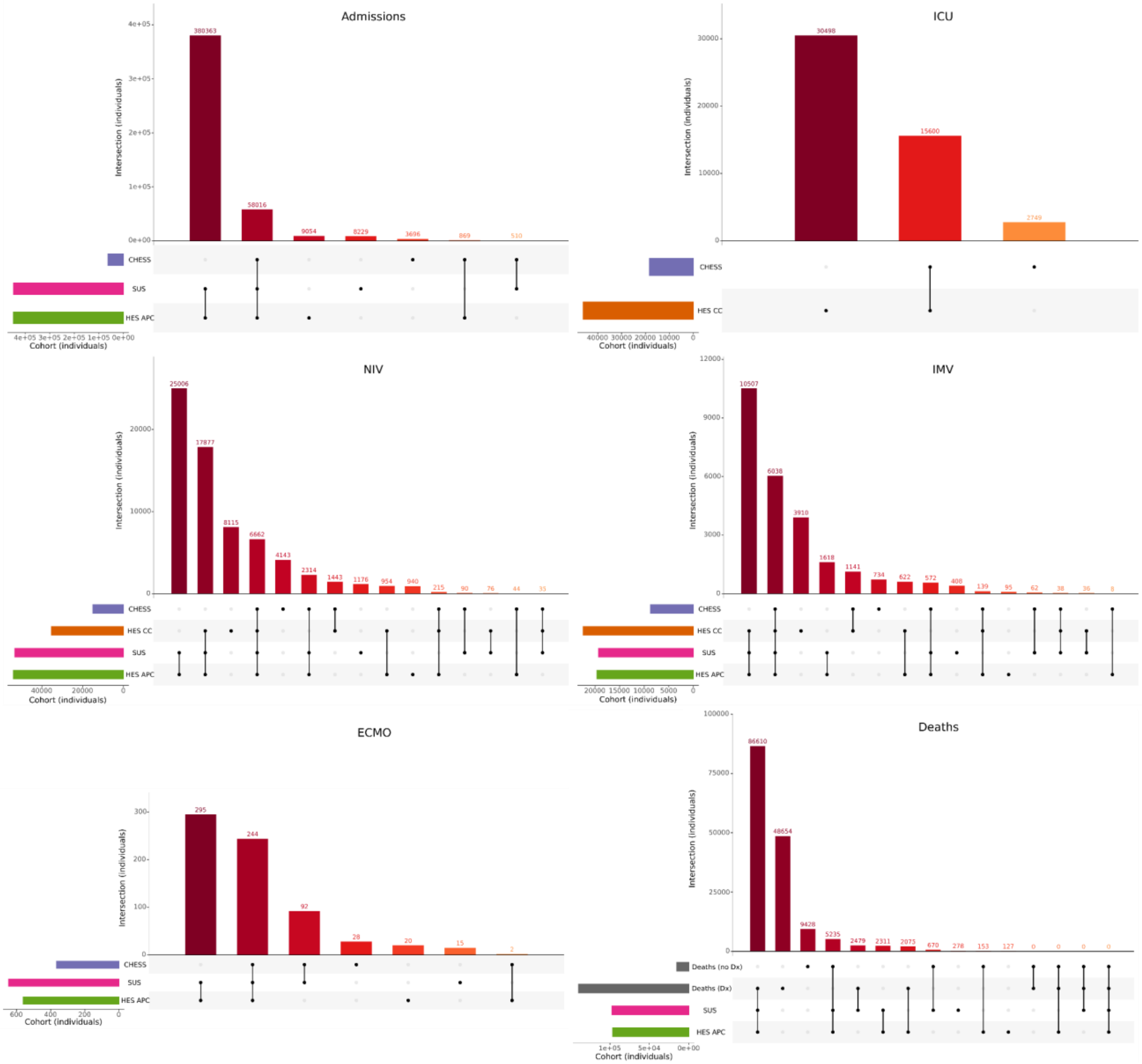
Ruth Gilbert	University College London
Ruth Norris	University of Manchester
Ruth Watkinson	University of Manchester
Ruwanthi Kolamunnage-Dona	University of Liverpool
Safa Salim	Imperial College London
Salil Deo	University of Glasgow
Sam Hollings	NHS Digital
Samantha Ip	University of Cambridge
Sandosh Padmanabhan	University of Glasgow
Sara Khalid	University of Oxford
Sarah Onida	Imperial College London
Sarah Steeg	University of Manchester
Seamus Kent	NICE
Seb Bacon	University of Oxford
Sebastian Vollmer	The Alan Turing Institute
Serban Stoica	University Hospital Bristol NHS Foundation Trust
Shane Johnson	Cancer Research UK
Sharmin Shabnam	University of Leicester
Shishir Rao	University of Oxford
Shubhra Sinha	University of Bristol
Sinduja Manohar	Health Data Research UK
Sonya Babu-Narayan	British Heart Foundation
Spencer Keene	University of Cambridge
Spiros Denaxas	University College London
Steve Ball	NHS Digital
Susheel Varma	Health Data Research UK
Tanja Mueller	University of Strathlcyde
Tapiwa Tungamirai	Wellcome Sanger Institute
Tasanee Braithwaite	Guy's and St Thomas' NHS Foundation Trust
Teri-Louise North	University of Bristol
Terry Quinn	University of Glasgow
Thomas Bolton	Health Data Research UK / BHF Data Science Centre
Thomas Lawrence	NICE
Tianxiao Wang	University of Cambridge

Tim Dong	University of Bristol
Tim Wilkinson	University of Edinburgh
Tom Norris	University of Leicester
Tom Palmer	University of Bristol
Tom Yates	University of Leicester
Tuankasfee Hama	University College London
Umesh Kadam	University of Leicester
Vahé Nafilyan	Office for National Statistics
Vandana Ayyar-Gupta	NICE
Vasa Curcin	King's College London
Venexia Walker	University of Bristol
William Whiteley	University of Edinburgh
Xiyun Jiang	University of Cambridge
Yat Yi Fan	University College London
Yi Mu	University College London
Yogini Chudasama	University of Leicester
Zainab Karim	British Heart Foundation

Supplementary Figures

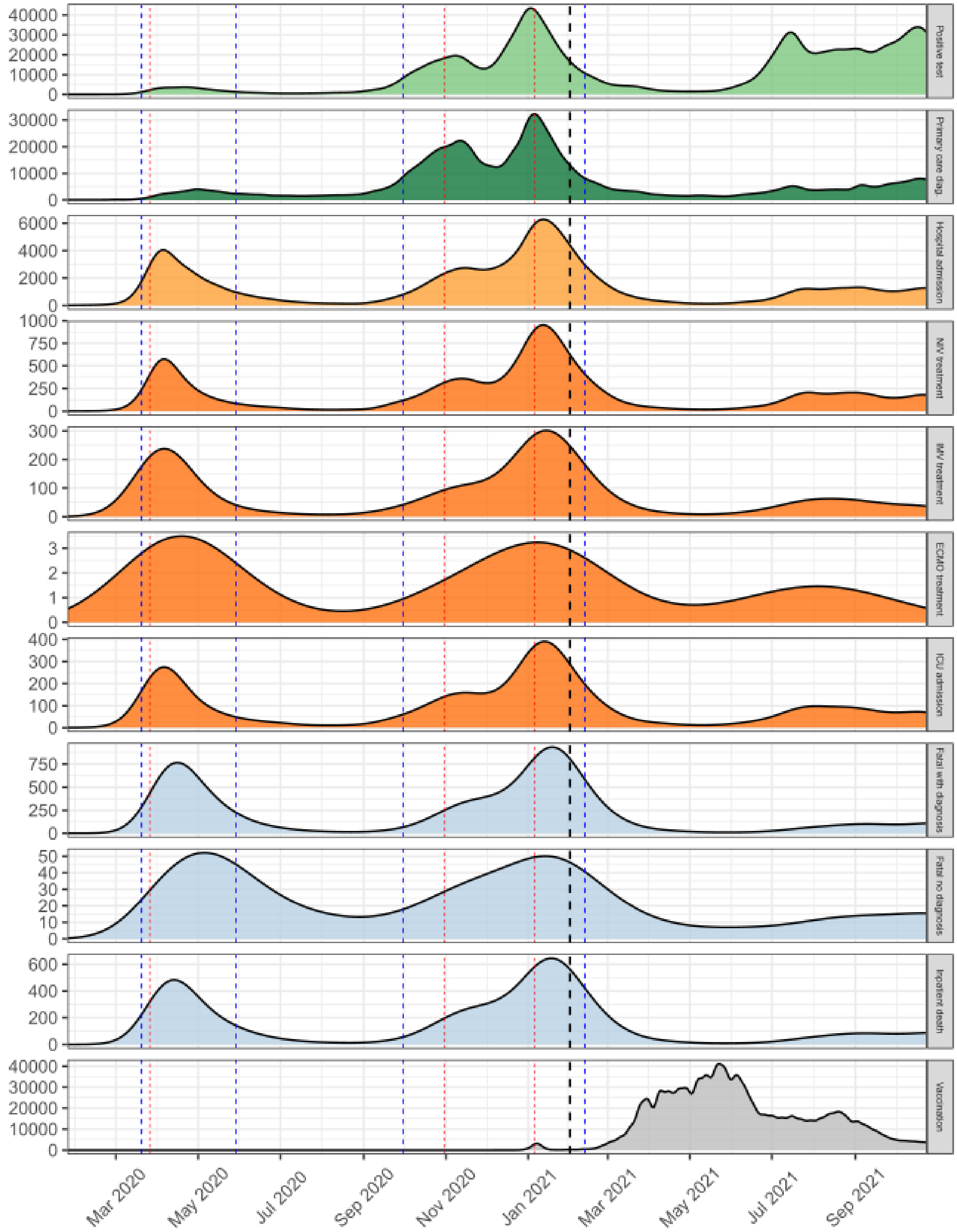

Individuals alive and registered in GDPPR on 23/01/2020 (n=57,032,174)
Individuals without a COVID-19 event (n=48,928,265)
Individuals with COVID-19 events (n=8,103,909)
Individuals with non-fatal events, without 28 days follow-up (n=858,984)
Final included individuals with COVID-19 events (n=7,244,925)

Supplementary figure 1: Flowchart of cohort design showing the number of records/individuals
Excluded individuals at different stages and the identification of cases and the final study population.

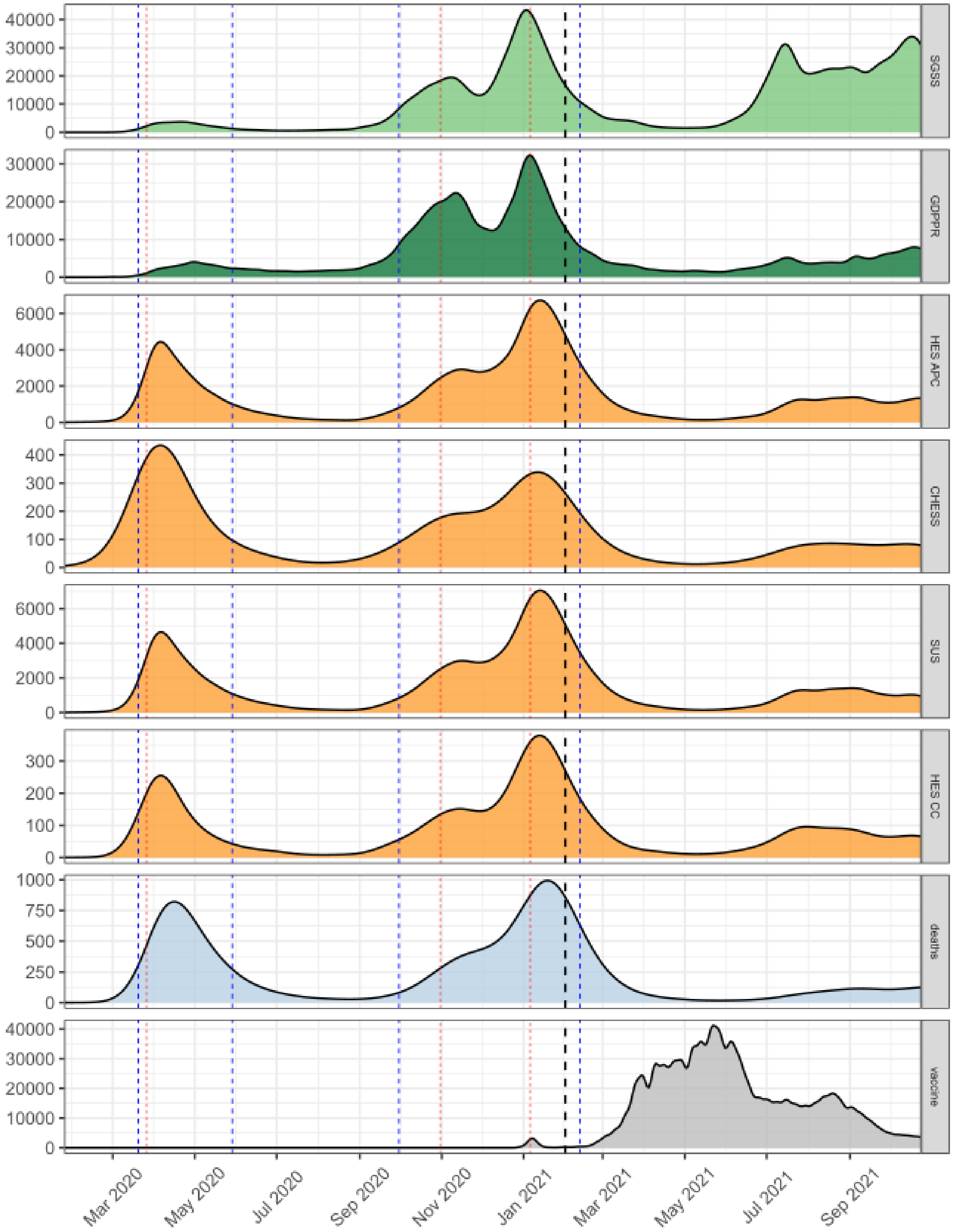


Supplementary figure 2: UpSet plots illustrating the numbers of individuals experiencing each COVID-19 event.

Vertical bars report unique individuals in the intersection denoted by the intersection matrix below. Horizontal bars report unique individuals identified from each dataset. Datasets are HES (Hospital episode statistics) APC (Admitted Patient Care) and CC (Critical Care), SUS (Secondary Uses Service) and CHESS (COVID-19 Hospitalisations in England Surveillance System) and ONS (Office of National Statistics) Civil Registration of Deaths. Positive tests and primary care diagnoses are not shown as these are derived from a single data source.

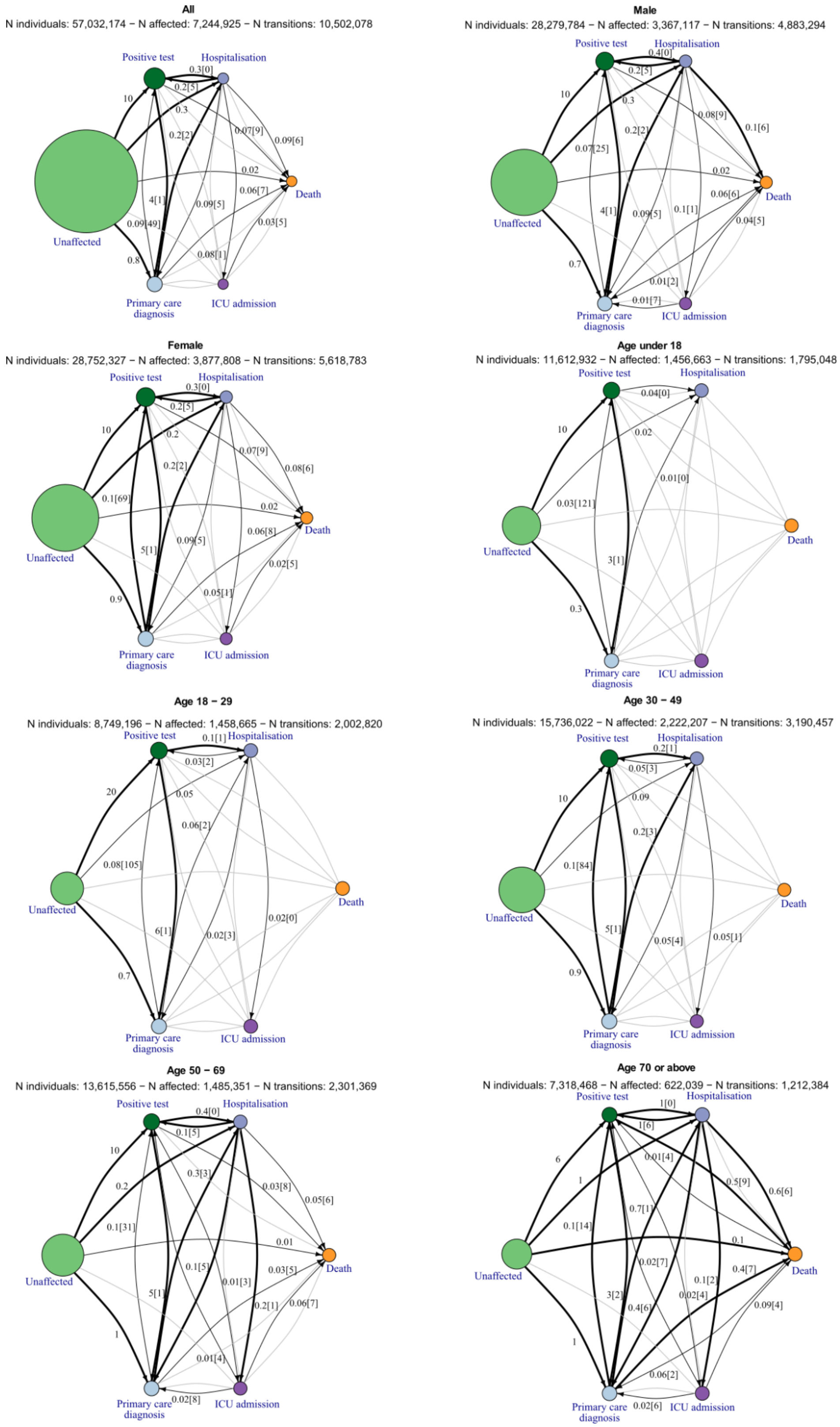


Supplementary figure 3: Timeline of COVID-19 event phenotypes.
Kernel density estimation plot showing unique events per individual per date, a person may have multiple events of the same type at different dates. Vaccination shows the date of the second dosage. Red vertical lines indicate the official English lockdown dates (26.03.2020, 31.10.2020 & 06.01.2021). Blue vertical lines indicate our study definition of wave 1 (20.03.2020 - 29.05.2020) and wave 2 (30.09.2020 - 12.02.2021). Black vertical line indicates the date used to explore the effects of vaccination on COVID-19 phenotypes.



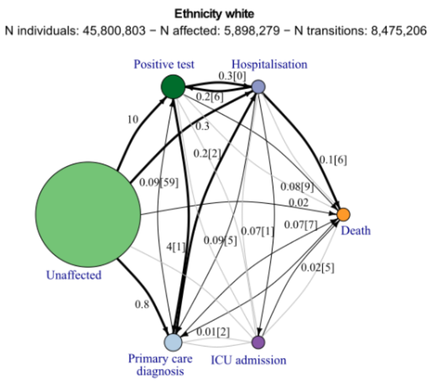
Supplementary figure 4: Timeline plots showing COVID-19 events, stratified by data source.
A person may have multiple events from the same source at different dates. Vaccination shows the date of the second dosage. Red vertical lines indicate the official English lockdown dates (26.03.2020, 31.10.2020 &

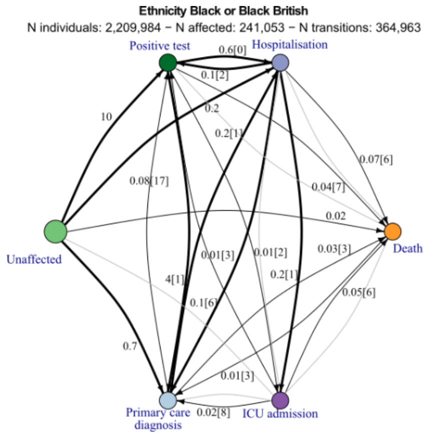
06.01.2021). Blue vertical lines indicate our study definition of wave 1 (20.03.2020 - 29.05.2020) and wave 2 (30.09.2020 - 12.02.2021). Black vertical line indicates the date used to explore the effects of vaccination on COVID-19 phenotypes.

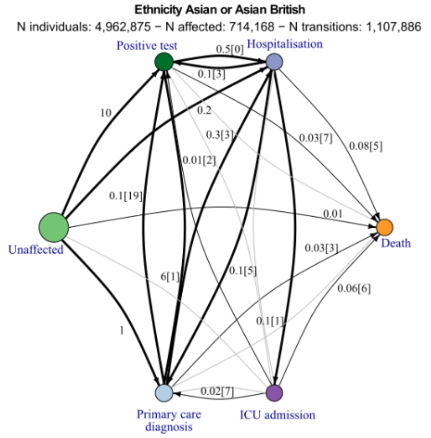


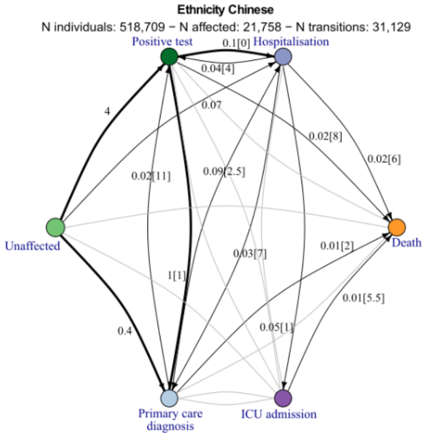
Supplementary figure 5: COVID-19 trajectory networks by gender and age groups.
Networks show percentage of individuals transitioning and the median number of days passing between severity phenotypes stratified on sex and age groups. The size of the circles represent the number of individuals with that

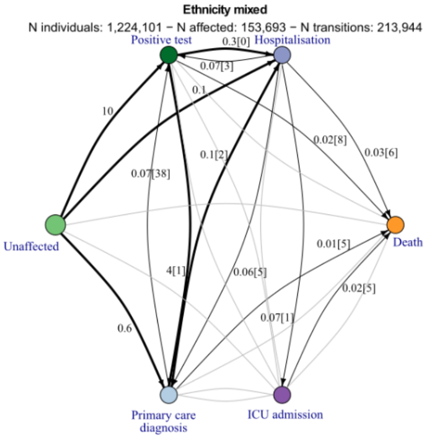
event relative to the total study population size of 57 million. Numbers on arrows are the percentage of individuals with the given transition (relative to N individuals in the group) and in square brackets median days between events across all individuals with that transition. Median days between unaffected and other severity phenotypes are larger, as these represent days from study start and the particular event. Thick arrows represent transitions occurring in ≥ 0.1%. Thin balck arrows represent transitions occurring in ≥ 0.01%. Any transitions occurring in fewer than 0.01% are not shown. N affected = N individuals with COVID-19.

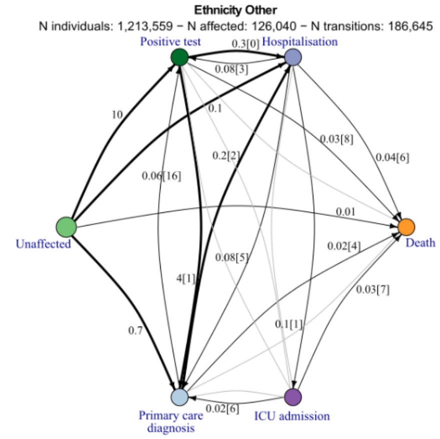

Ethnicity white
N individuals: 45,800,803 − N affected: 5,898,279 − N transitions: 8,475,206
Positive test
Hospitalisation
0.3[0]
0.2[6]
10
0.3
0.2[2]
0.1[6]
0.09[59]
0.08[9]
0.02
0.07[7]
Death
0.07[1]
4[1]
0.09[5]
0.02[5]
Unaffected
0.8
0.01[2]
Primary care diagnosis
ICU admission

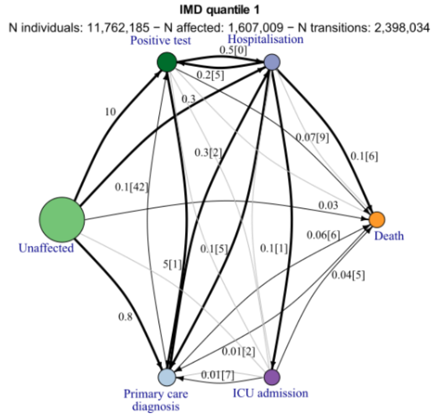

Ethnicity Black or Black British
N individuals: 2,209,984 − N affected: 241,053 − N transitions: 364,963
Positive test
0.6[0]
Hospitalisation
0.1[2]
10
0.2
0.2[1]
0.07[6]
0.08[17]
0.04[7]
0.02
Unaffected
0.01[3]
0.01[2]
0.03[3]
Death
4[1]
0.2[1]
0.05[6]
0.1[6]
0.7
0.01[3]
Primary care diagnosis
0.02[8]
ICU admission

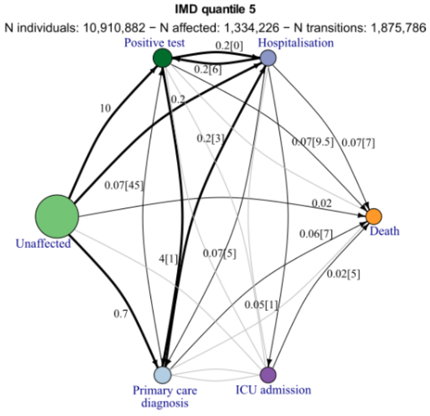

Ethnicity Asian or Asian British
N individuals: 4,962,875 − N affected: 714,168 − N transitions: 1,107,886
Positive test
0.5[0]
Hospitalisation
0.1[3]
10
0.2
0.3[3]
0.03[7]
0.08[5]
0.01[2]
0.1[19]
0.01
Unaffected
0.03[3]
Death
6[1]
0.1[5]
0.06[6]
0.1[1]
1
0.02[7]
Primary care diagnosis
ICU admission


Ethnicity Chinese
N individuals: 518,709 − N affected: 21,758 − N transitions: 31,129
Positive test
0.1[0]
Hospitalisation
0.04[4]
4
0.07
0.02[8]
0.02[6]
0.02[11]
0.09[2.5]
Unaffected
Death
0.03[7]
0.01[2]
1[1]
0.01[5.5]
0.05[1]
0.4
Primary care diagnosis
ICU admission

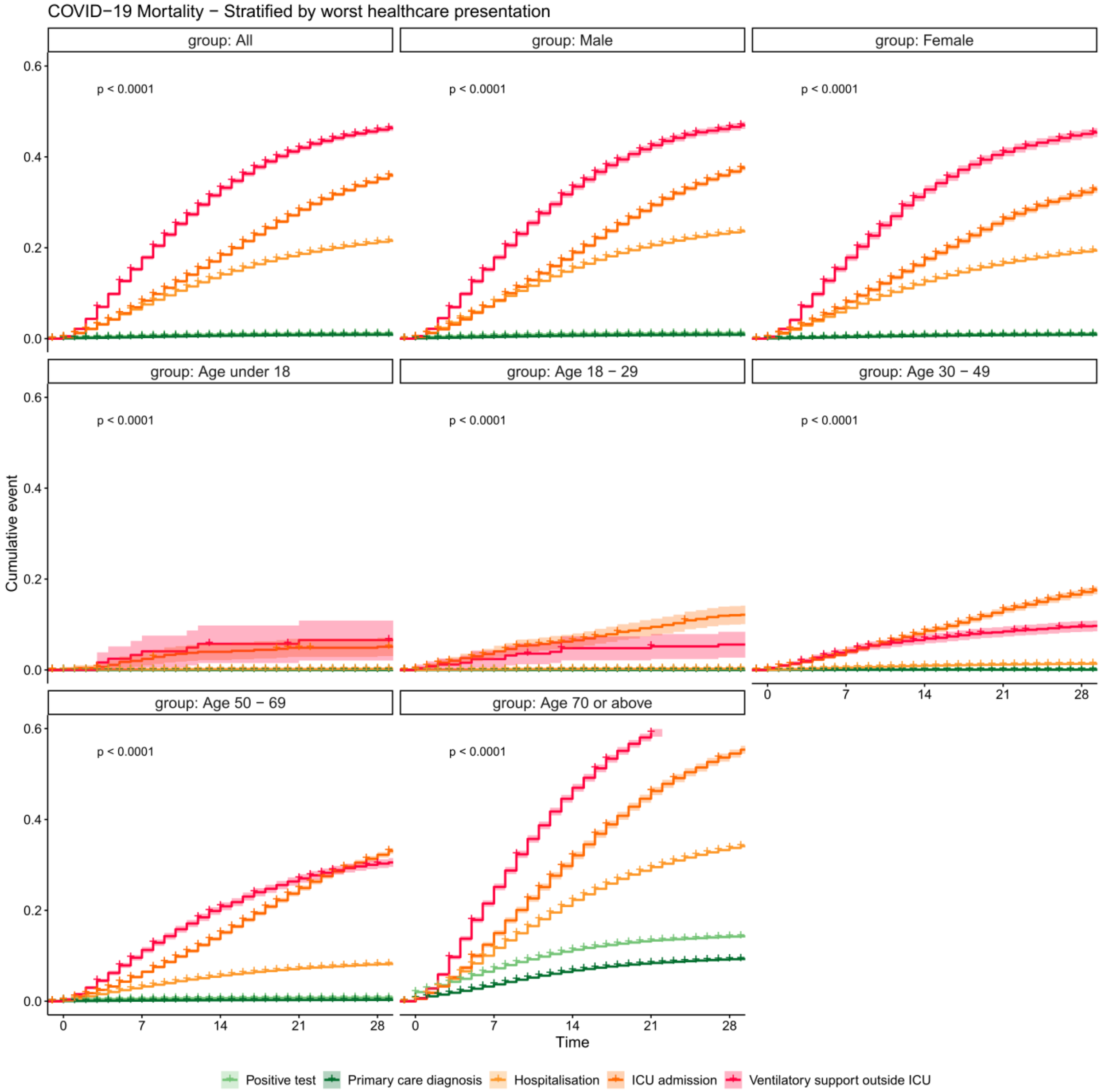

Ethnicity mixed
N individuals: 1,224,101 − N affected: 153,693 − N transitions: 213,944
Positive test
0.3[0]
Hospitalisation
0.07[3]
0.1
10
0.02[8]
0.03[6]
0.1[2]
0.07[38]
Unaffected
0.01[5]
Death
4[1]
0.06[5]
0.02[5]
0.07[1]
0.6
Primary care diagnosis
ICU admission


Ethnicity Other
N individuals: 1,213,559 − N affected: 126,040 − N transitions: 186,645
Positive test
0.3[0]
Hospitalisation
0.08[3]
10
0.1
0.03[8]
0.04[6]
0.2[2]
0.06[16]
0.01
Unaffected
Death
0.02[4]
0.08[5]
4[1]
0.03[7]
0.1[1]
0.7
0.02[6]
Primary care diagnosis
ICU admission


IMD quantile 1
N individuals: 11,762,185 − N affected: 1,607,009 − N transitions: 2,398,034
Positive test
Hospitalisation
0.5[0]
0.2[5]
10
0.3
0.07[9]
0.3[2]
0.1[6]
0.1[42]
0.03
Unaffected
0.06[6]
Death
0.1[5]
0.1[1]
5[1]
0.04[5]
0.8
0.01[2]
0.01[7]
Primary care diagnosis
ICU admission

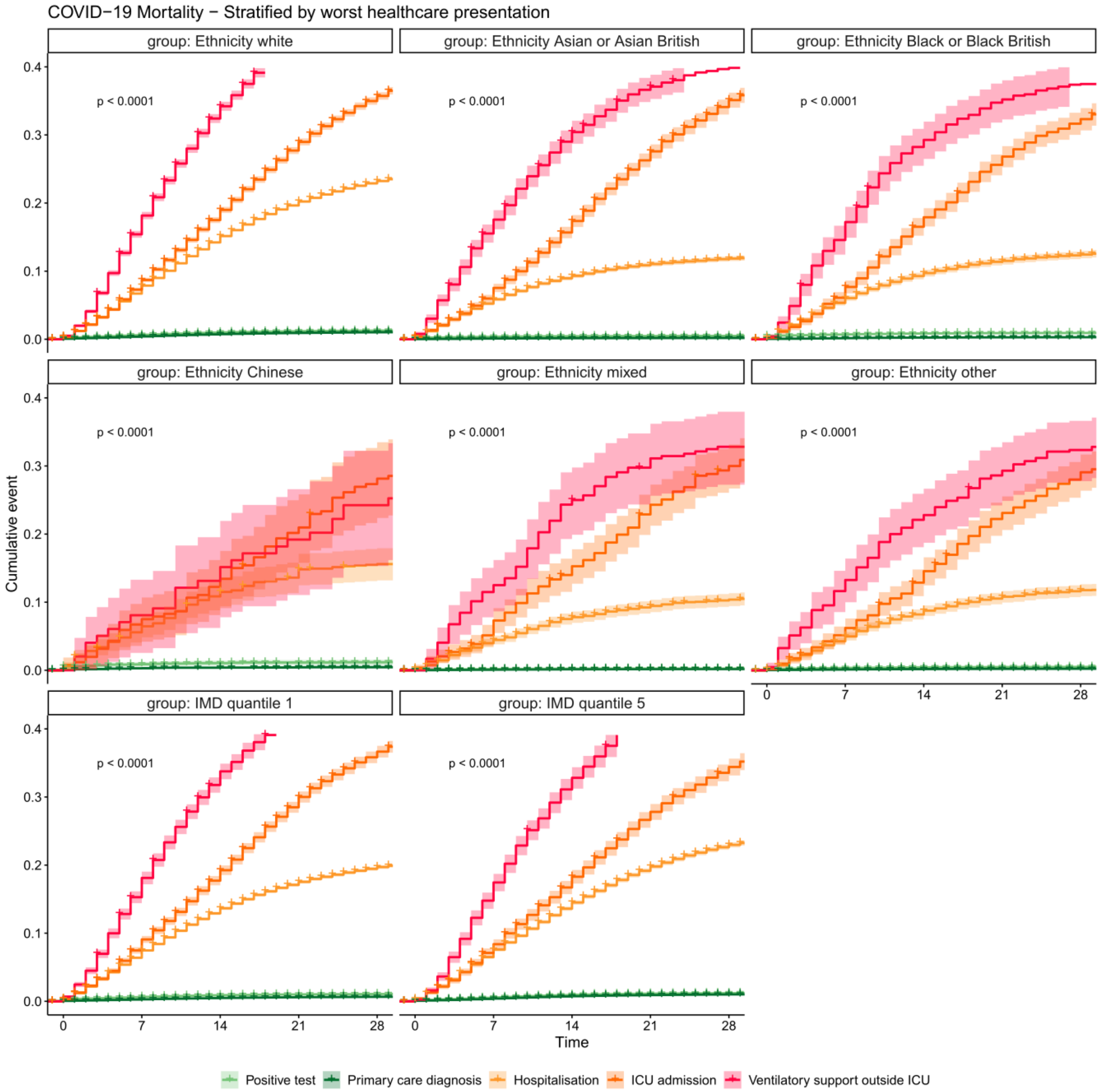

IMD quantile 5
N individuals: 10,910,882 − N affected: 1,334,226 − N transitions: 1,875,786
Positive test
0.2[0]
Hospitalisation
0.2[6]
10
0.2
0.2[3]
0.07[9.5]
0.07[7]
0.07[45]
0.02
Unaffected
0.06[7]
Death
4[1]
0.07[5]
0.02[5]
0.05[1]
0.7
Primary care diagnosis
ICU admission

Supplementary figure 6: COVID-19 trajectory networks by ethnicity and IMD.
Networks show showing percentage of individuals transitioning and the median number of days passing between severity phenotypes stratified on sex and age groups. The size of the circles represent the number of individuals with that event relative to the total study population size of 57 million. Numbers on arrows are the percentage of individuals with the given transition (relative to N individuals in the group) and in square brackets median days between events across all individuals with that transition. Median days between unaffected and other severity phenotypes are larger, as these represent days from study start and the particular event. Thick arrows represent transitions occurring in ≥ 0.1%. Thin balck arrows represent transitions occurring in ≥ 0.01%. Any transitions occurring in fewer than 0.01% are not shown. N affected = N individuals with COVID-19.



Supplementary figure 7: Kaplan Meier plot of COVID-19 mortality by gender and age groups.
Curves are stratified by worst healthcare presentation, here listed in increasing order of severity; positive test (light green), primary care diagnosis (dark green), hospitalisation (yellow), ICU admission (orange) and ventilatory

support outside ICU (red). Note the ICU admission group does not include patients who received ventilatory support outside of ICU wards. Shaded areas represent 95% confidence intervals in all panels.



Supplementary figure 8: Kaplan Meier plot of COVID-19 mortality by ethnicity and IMD level. Curves are stratified by worst healthcare presentation, here listed in increasing order of severity; positive test (light green), primary care diagnosis (dark green), hospitalisation (yellow), ICU admission (orange) and ventilatory support outside ICU (red). Note the ICU admission group does not include patients who received ventilatory support outside of ICU wards. Shaded areas represent 95% confidence intervals in all panels.

Supplementary Tables

Supplementary table 1: Number of individuals identified from each data source stratified by COVID-19 event and as total individuals across all data sources.

Date ranges shown for all included data sources after filtering as per cohort definition (see Figure 1). "Death without diagnosis" refers to patients that died within 28 days of a COVID-related event where COVID-19 was not the recorded cause of death on the death certificate.

Data	SGSS	GDPPR	HES APC	SUS	CHESS	HES CC	Deaths	Total
Min. Date	2020-01-24	2020-01-23	2020-01-23	2020-01-23	2020-01-23	2020-01-28	2020-01-30	2020-01-23
Max. Date	2021-11-30	2021-11-30	2021-11-30	2021-11-24	2021-11-30	2021-11-30	2021-11-30	2021-11-30
Positive test	6,778,342							6,778,342
GP diagnosis		3,056,132						3,056,132
Hospitalisation			448,302	447,118	63,091			460,737
ECMO treatment			561	646	366			696
ICU admission					18,349	46,098		48,847
IMV treatment			19,599	19,279	8,732	22,431		25,928
NIV treatment			54,012	53,236	14,946	35,377		69,090
Inpatient Death			96,511	97,583				99,938
Fatal with diagnosis							139,818	139,818
Death without diagnosis							15,486	15,486

Supplementary table 2: COVID-19 codes & frequencies.

Codes appearing with a frequency <5 are masked as '<5' to protect privacy.

COVID-19 Event	Code	Terminology	Description	Source	n
01_Covid_positive_test				SGSS	7135843
01_GP_covid_diagnosis	1240581000000104	SNOMED	Severe acute respiratory syndrome coronavirus 2 ribonucleic acid detected (finding)	GDPPR	2304566
01_GP_covid_diagnosis	1300721000000109	SNOMED	Coronavirus disease 19 caused by severe acute respiratory syndrome coronavirus 2 confirmed by laboratory test (situation)	GDPPR	718451
01_GP_covid_diagnosis	1008541000000105	SNOMED	Coronavirus ribonucleic acid detection assay (observable entity)	GDPPR	435378
01_GP_covid_diagnosis	1321541000000108	SNOMED	Severe acute respiratory syndrome coronavirus 2 immunoglobulin G detected (finding)	GDPPR	330333
01_GP_covid_diagnosis	1240751000000100	SNOMED	Coronavirus disease 19 caused by severe acute respiratory syndrome coronavirus 2 (disorder)	GDPPR	195948
01_GP_covid_diagnosis	1322781000000102	SNOMED	Severe acute respiratory syndrome coronavirus 2 antigen detection result positive (finding)	GDPPR	96279
01_GP_covid_diagnosis	186747009	SNOMED	Coronavirus infection (disorder)	GDPPR	61017
01_GP_covid_diagnosis	1300681000000102	SNOMED	Assessment using coronavirus disease 19 severity scale (procedure)	GDPPR	38523
01_GP_covid_diagnosis	1300731000000106	SNOMED	Coronavirus disease 19 caused by severe acute respiratory syndrome coronavirus 2 confirmed using clinical diagnostic criteria (situation)	GDPPR	34344
01_GP_covid_diagnosis	1240511000000106	SNOMED	Detection of severe acute respiratory syndrome coronavirus 2 using polymerase chain reaction technique (procedure)	GDPPR	25087
01_GP_covid_diagnosis	1240741000000103	SNOMED	Severe acute respiratory syndrome coronavirus 2 serology (observable entity)	GDPPR	14850
01_GP_covid_diagnosis	1240551000000105	SNOMED	Pneumonia caused by severe acute respiratory syndrome coronavirus 2 (disorder)	GDPPR	13221
01_GP_covid_diagnosis	1321761000000103	SNOMED	Severe acute respiratory syndrome coronavirus 2 immunoglobulin A detected (finding)	GDPPR	8269

01_GP_covid_diagnosis	1322871000000109	SNOMED	Severe acute respiratory syndrome coronavirus 2 antibody detection result positive (finding)	GDPPR	6683
01_GP_covid_diagnosis	1300631000000101	SNOMED	Coronavirus disease 19 severity score (observable entity)	GDPPR	6179
01_GP_covid_diagnosis	1240541000000107	SNOMED	Infection of upper respiratory tract caused by severe acute respiratory syndrome coronavirus 2 (disorder)	GDPPR	4568
01_GP_covid_diagnosis	1300671000000104	SNOMED	Coronavirus disease 19 severity scale (assessment scale)	GDPPR	2822
01_GP_covid_diagnosis	1029481000000103	SNOMED	Coronavirus nucleic acid detection assay (observable entity)	GDPPR	2737
01_GP_covid_diagnosis	1321551000000106	SNOMED	Severe acute respiratory syndrome coronavirus 2 immunoglobulin M detected (finding)	GDPPR	2423
01_GP_covid_diagnosis	1240401000000105	SNOMED	Antibody to severe acute respiratory syndrome coronavirus 2 (substance)	GDPPR	634
01_GP_covid_diagnosis	1240571000000101	SNOMED	Gastroenteritis caused by severe acute respiratory syndrome coronavirus 2 (disorder)	GDPPR	350
01_GP_covid_diagnosis	1240421000000101	SNOMED	Serotype severe acute respiratory syndrome coronavirus 2 (qualifier value)	GDPPR	298
01_GP_covid_diagnosis	1240531000000103	SNOMED	Myocarditis caused by severe acute respiratory syndrome coronavirus 2 (disorder)	GDPPR	236
01_GP_covid_diagnosis	1321341000000103	SNOMED	Arbitrary concentration of severe acute respiratory syndrome coronavirus 2 immunoglobulin G in serum (observable entity)	GDPPR	208
01_GP_covid_diagnosis	1240391000000107	SNOMED	Antigen of severe acute respiratory syndrome coronavirus 2 (substance)	GDPPR	163
01_GP_covid_diagnosis	1321331000000107	SNOMED	Arbitrary concentration of severe acute respiratory syndrome coronavirus 2 total immunoglobulin in serum (observable entity)	GDPPR	134
01_GP_covid_diagnosis	1321301000000101	SNOMED	Severe acute respiratory syndrome coronavirus 2 ribonucleic acid qualitative existence in specimen (observable entity)	GDPPR	122
01_GP_covid_diagnosis	1240561000000108	SNOMED	Encephalopathy caused by severe acute respiratory syndrome coronavirus 2 (disorder)	GDPPR	83
01_GP_covid_diagnosis	1321351000000100	SNOMED	Arbitrary concentration of severe acute respiratory syndrome coronavirus 2 immunoglobulin M in serum (observable entity)	GDPPR	25
01_GP_covid_diagnosis	1321241000000105	SNOMED	Cardiomyopathy caused by severe acute respiratory syndrome coronavirus 2 (disorder)	GDPPR	23
01_GP_covid_diagnosis	1240521000000100	SNOMED	Otitis media caused by severe acute respiratory syndrome coronavirus 2 (disorder)	GDPPR	21
01_GP_covid_diagnosis	1321321000000105	SNOMED	Severe acute respiratory syndrome coronavirus 2 immunoglobulin G qualitative existence in specimen (observable entity)	GDPPR	16
01_GP_covid_diagnosis	1321311000000104	SNOMED	Severe acute respiratory syndrome coronavirus 2 immunoglobulin M qualitative existence in specimen (observable entity)	GDPPR	11
01_GP_covid_diagnosis	1240411000000107	SNOMED	Ribonucleic acid of severe acute respiratory syndrome coronavirus 2 (substance)	GDPPR	8
01_GP_covid_diagnosis	120814005	SNOMED	Coronavirus antibody (substance)	GDPPR	<5
01_GP_covid_diagnosis	1321811000000105	SNOMED	Severe acute respiratory syndrome coronavirus 2 immunoglobulin A qualitative existence in specimen (observable entity)	GDPPR	<5
01_GP_covid_diagnosis	1321201000000107	SNOMED	Coronavirus disease 19 caused by severe acute respiratory syndrome coronavirus 2 health issues simple reference set (foundation metadata concept)	GDPPR	<5
01_GP_covid_diagnosis	1321181000000108	SNOMED	Coronavirus disease 19 caused by severe acute respiratory syndrome coronavirus 2 record extraction simple reference set (foundation metadata concept)	GDPPR	<5
01_GP_covid_diagnosis	1240381000000105	SNOMED	Severe acute respiratory syndrome coronavirus 2 (organism)	GDPPR	<5
01_GP_covid_diagnosis	1321801000000108	SNOMED	Arbitrary concentration of severe acute respiratory syndrome coronavirus 2 immunoglobulin A in serum (observable entity)	GDPPR	<5

01_GP_covid_diagnosis	1321191000000105	SNOMED	Coronavirus disease 19 caused by severe acute respiratory syndrome coronavirus 2 procedures simple reference set (foundation metadata concept)	GDPPR	<5
02_Covid_admission	U07.1	ICD10	Confirmed_COVID19	HES APC	826227
02_Covid_admission	U07.1	ICD10	Confirmed_COVID19	SUS	820696
02_Covid_admission	U07.2	ICD10	Suspected_COVID19	SUS	69207
02_Covid_admission			HospitalAdmissionDate IS NOT null	CHESS	67029
02_Covid_admission	U07.2	ICD10	Suspected_COVID19	HES APC	64124
03_ECMO_treatment	X58.1	OPCS	Extracorporeal membrane oxygenation	SUS	736
03_ECMO_treatment	X58.1	OPCS	Extracorporeal membrane oxygenation	HES APC	629
03_ECMO_treatment			RespiratorySupportECMO == Yes	CHESS	439
03_ICU_admission			id is in hes_cc table	HES CC	57338
03_ICU_admission			DateAdmittedICU IS NOT null	CHESS	19629
03_IMV_treatment			ARESSUPDAYS > 0	HES CC	27555
03_IMV_treatment	E85.1	OPCS	Invasive ventilation	SUS	21737
03_IMV_treatment	E85.1	OPCS	Invasive ventilation	HES APC	21696
03_IMV_treatment			Invasivemechanicalventilation == Yes	CHESS	9572
03_IMV_treatment	X56	OPCS	Intubation of trachea	SUS	520
03_IMV_treatment	X56	OPCS	Intubation of trachea	HES APC	498
03_NIV_treatment	E85.6	OPCS	Continuous positive airway pressure	HES APC	47095
03_NIV_treatment	E85.6	OPCS	Continuous positive airway pressure	SUS	46150
03_NIV_treatment			bressupdays > 0	HES CC	40273
03_NIV_treatment	E85.2	OPCS	Non-invasive ventilation NEC	SUS	20912
03_NIV_treatment	E85.2	OPCS	Non-invasive ventilation NEC	HES APC	20052
03_NIV_treatment			Highflownasaloxygen OR NoninvasiveMechanicalventilation == Yes	CHESS	15243
04_Covid_inpatient_death			DISCHARGE_METHOD_HOSPITAL_PROVIDER_SPELL = 4 (Died)	SUS	99247
04_Covid_inpatient_death			DISMETH = 4 (Died)	HES APC	96481
04_Covid_inpatient_death			DISCHARGE_DESTINATION_HOSPITAL_PROVIDER_SPELL = 79 (Not applicable - PATIENT died or still birth)	SUS	169
04_Covid_inpatient_death			DISDEST = 79 (Not applicable - PATIENT died or still birth)	HES APC	122
04_Fatal_with_covid_diagnosis	U071	ICD10		deaths	137636
04_Fatal_with_covid_diagnosis	U072	ICD10		deaths	4052
04_Fatal_without_covid_diagnosis			ONS death within 28 days	deaths	15489

Supplementary table 3: 270 CALIBER phenotypes, aggregated into 16 categories
Table shows the number of individuals within the study cohort identified from GDPPR (SNOMED-CT) and HES APC (ICD-10, OPCS-4).

Category	Phenotype	Individuals
Benign neoplasm/CIN	Benign neoplasm of colon rectum anus and anal canal	132,843
Benign neoplasm/CIN	Benign neoplasm of ovary	95,700
Benign neoplasm/CIN	Benign neoplasm and polyp of uterus	49,834
Benign neoplasm/CIN	Benign neoplasm of stomach and duodenum	37,515
Benign neoplasm/CIN	Haemangioma any site	20,943
Benign neoplasm/CIN	Carcinoma in situ cervical	8,605
Benign neoplasm/CIN	Benign neoplasm of brain and other parts of central nervous system	4,255
Cancers	Myelodysplastic syndromes	534,599
Cancers	Primary malignancy other organs	439,688
Cancers	Primary malignancy other skin and subcutaneous tissue	423,906
Cancers	Primary malignancy breast	174,836
Cancers	Primary malignancy cervical	85,510
Cancers	Primary malignancy prostate	28,726
Cancers	Primary malignancy colorectal and anus	20,521
Cancers	Primary malignancy malignant melanoma	16,639
Cancers	Non-hodgkin lymphoma	12,737
Cancers	Secondary malignancy other organs	10,425
Cancers	Leukaemia	9,904
Cancers	Primary malignancy bladder	9,738
Cancers	Monoclonal gammopathy of undetermined significance	9,229
Cancers	Hodgkin lymphoma	9,070
Cancers	Multiple myeloma and malignant plasma cell neoplasms	7,521
Cancers	Primary malignancy lung and trachea	5,796
Cancers	Secondary malignancy bone	5,510
Cancers	Primary malignancy kidney and ureter	5,134
Cancers	Primary malignancy testicular	4,684
Cancers	Primary malignancy uterine	4,249
Cancers	Secondary malignancy lung	3,979
Cancers	Secondary malignancy liver and intrahepatic bile duct	3,618
Cancers	Primary malignancy ovarian	3,520
Cancers	Primary malignancy oro-pharyngeal	3,297
Cancers	Primary malignancy thyroid	3,144
Cancers	Polycythaemia vera	3,138
Cancers	Secondary malignancy retroperitoneum and peritoneum	2,140
Cancers	Primary malignancy brain other CNS and intracranial	1,609
Cancers	Primary malignancy oesophageal	1,412
Cancers	Secondary malignancy lymph nodes	1,255
Cancers	Primary malignancy liver	1,071
Cancers	Secondary malignancy brain other CNS and intracranial	991
Cancers	Primary malignancy stomach	965
Cancers	Secondary malignancy pleura	722
Cancers	Primary malignancy bone and articular cartilage	671
Cancers	Primary malignancy pancreatic	653
Cancers	Secondary malignancy bowel	528
Cancers	Primary malignancy biliary tract	397
Cancers	Secondary malignancy adrenal gland	382
Cancers	Primary malignancy mesothelioma	88
Cancers	Primary malignancy multiple independent sites	9
Diseases of the circulatory system	Hypertension	856,557

Diseases of the circulatory system	Coronary heart disease not otherwise specified	224,719
Diseases of the circulatory system	Stable angina	152,989
Diseases of the circulatory system	Atrial fibrillation	139,086
Diseases of the circulatory system	Myocardial infarction	122,024
Diseases of the circulatory system	Heart failure	77,189
Diseases of the circulatory system	Stroke NOS	65,937
Diseases of the circulatory system	Transient ischaemic attack	64,896
Diseases of the circulatory system	Ischaemic stroke	59,096
Diseases of the circulatory system	Peripheral arterial disease	57,162
Diseases of the circulatory system	Unstable angina	56,033
Diseases of the circulatory system	Supraventricular tachycardia	43,127
Diseases of the circulatory system	Venous thromboembolic disease excluding PE	34,362
Diseases of the circulatory system	Right bundle branch block	32,120
Diseases of the circulatory system	Left bundle branch block	26,332
Diseases of the circulatory system	Atrioventricular block first degree	22,559
Diseases of the circulatory system	Abdominal aortic aneurysm	17,123
Diseases of the circulatory system	Raynaud's syndrome	13,716
Diseases of the circulatory system	Other cardiomyopathy	10,541
Diseases of the circulatory system	Pericardial effusion noninflammatory	10,040
Diseases of the circulatory system	Secondary pulmonary hypertension	9,934
Diseases of the circulatory system	Atrioventricular block complete	9,216
Diseases of the circulatory system	Ventricular tachycardia	8,598
Diseases of the circulatory system	Intracerebral haemorrhage	8,015
Diseases of the circulatory system	Dilated cardiomyopathy	7,001
Diseases of the circulatory system	Atrioventricular block second degree	6,784
Diseases of the circulatory system	Sick sinus syndrome	5,579
Diseases of the circulatory system	Primary pulmonary hypertension	4,457
Diseases of the circulatory system	Subdural haematoma - nontraumatic	4,188
Diseases of the circulatory system	Hypertrophic cardiomyopathy	2,858
Diseases of the circulatory system	Bifascicular block	2,441
Diseases of the circulatory system	Trifascicular block	2,438
Diseases of the circulatory system	Subarachnoid haemorrhage	1,094
Diseases of the circulatory system	Pulmonary embolism	616
Diseases of the circulatory system	Rheumatic valve disease	206
Diseases of the digestive system	Abdominal hernia	208,656
Diseases of the digestive system	Appendicitis	126,466
Diseases of the digestive system	Oesophagitis and oesophageal ulcer	113,970
Diseases of the digestive system	Cholecystitis	86,338
Diseases of the digestive system	Fatty liver	43,473
Diseases of the digestive system	Anal fissure	32,084
Diseases of the digestive system	Coeliac disease	30,898
Diseases of the digestive system	Peritonitis	29,371
Diseases of the digestive system	Barretts oesophagus	26,380
Diseases of the digestive system	Anorectal fistula	21,208
Diseases of the digestive system	Liver fibrosis sclerosis and cirrhosis	17,773
Diseases of the digestive system	Pancreatitis	10,946
Diseases of the digestive system	Anorectal prolapse	10,900
Diseases of the digestive system	Cholangitis	10,375
Diseases of the digestive system	Gastro-oesophageal reflux disease	8,512
Diseases of the digestive system	Portal hypertension	7,961
Diseases of the digestive system	Volvulus	6,271
Diseases of the digestive system	Autoimmune liver disease	4,763
Diseases of the digestive system	Angiodysplasia of colon	4,507
Diseases of the digestive system	Oesophageal varices	3,868
Diseases of the digestive system	Diaphragmatic hernia	3,464
Diseases of the digestive system	Alcoholic liver disease	2,988
Diseases of the digestive system	Peptic ulcer disease	2,038
Diseases of the digestive system	Hepatic failure	2,031
Diseases of the digestive system	Diverticular disease of intestine acute and chronic	316
Diseases of the ear	Hearing loss	8,566
Diseases of the ear	Tinnitus	7,785
Diseases of the ear	Meniere disease	5,628
Diseases of the endocrine system	Diabetes	410,730

Diseases of the endocrine system	Obesity	344,462
Diseases of the endocrine system	Hypo or hyperthyroidism	272,744
Diseases of the endocrine system	Polycystic ovarian syndrome	44,632
Diseases of the endocrine system	Syndrome of inappropriate secretion of antidiuretic hormone	31,000
Diseases of the endocrine system	Hyperparathyroidism	10,251
Diseases of the endocrine system	Cystic fibrosis	1,829
Diseases of the eye	Diabetic ophthalmic complications	295,315
Diseases of the eye	Cataract	236,092
Diseases of the eye	Glaucoma	45,545
Diseases of the eye	Macular degeneration	41,547
Diseases of the eye	Retinal detachments and breaks	25,472
Diseases of the eye	Ptosis of eyelid	12,169
Diseases of the eye	Anterior and intermediate uveitis	5,325
Diseases of the eye	Keratitis	901
Diseases of the eye	Scleritis and episcleritis	703
Diseases of the eye	Posterior uveitis	455
Diseases of the eye	Retinal vascular occlusions	167
Diseases of the eye	Visual impairment and blindness	79
Diseases of the genitourinary system	Menorrhagia and polymenorrhoea	145,324
Diseases of the genitourinary system	Acute kidney injury	134,152
Diseases of the genitourinary system	Urolithiasis	94,047
Diseases of the genitourinary system	Obstructive and reflux uropathy	48,394
Diseases of the genitourinary system	Urinary incontinence	47,483
Diseases of the genitourinary system	Postmenopausal bleeding	34,839
Diseases of the genitourinary system	Hydrocoele including infected	32,140
Diseases of the genitourinary system	Dysmenorrhoea	29,075
Diseases of the genitourinary system	Glomerulonephritis	24,413
Diseases of the genitourinary system	End stage renal disease	23,596
Diseases of the genitourinary system	Postcoital and contact bleeding	22,840
Diseases of the genitourinary system	Endometrial hyperplasia and hypertrophy	21,708
Diseases of the genitourinary system	Non-acute cystitis	8,645
Diseases of the genitourinary system	Erectile dysfunction	4,017
Diseases of the genitourinary system	Tubulo-interstitial nephritis	2,149
Diseases of the genitourinary system	Undescended testicle	72
Diseases of the genitourinary system	Male infertility	30
Diseases of the respiratory system	Asthma	1,078,007
Diseases of the respiratory system	Allergic and chronic rhinitis	137,547
Diseases of the respiratory system	COPD	122,793
Diseases of the respiratory system	Chronic sinusitis	113,255
Diseases of the respiratory system	Sleep apnoea	76,997
Diseases of the respiratory system	Pulmonary collapse excluding pneumothorax	43,196
Diseases of the respiratory system	Hypertrophy of nasal turbinates	34,629
Diseases of the respiratory system	Bronchiectasis	26,741
Diseases of the respiratory system	Aspiration pneumonitis	16,550
Diseases of the respiratory system	Other interstitial pulmonary diseases with fibrosis	14,342
Diseases of the respiratory system	Pleural plaque	8,299
Diseases of the respiratory system	Respiratory failure	2,580
Diseases of the respiratory system	Asbestosis	1,108
Diseases of the respiratory system	Pleural effusion	117
Diseases of the respiratory system	Pneumothorax	46
Diseases of the respiratory system	Nasal polyp	0
Haematological immunological conditions	Secondary or other thrombocytopaenia	22,495
Haematological immunological conditions	Splenomegaly	16,652
Haematological immunological conditions	Sickle-cell trait	10,840
Haematological immunological conditions	Thalassaemia trait	9,113
Haematological immunological conditions	Primary or idiopathic thrombocytopaenia	7,652
Haematological immunological conditions	Hyposplenism	7,470
Haematological immunological conditions	Thrombophilia	6,333
Haematological immunological conditions	Thalassaemia	5,678
Haematological immunological conditions	Agranulocytosis	5,375
Haematological immunological conditions	Sickle-cell anaemia	4,718
Haematological immunological conditions	Secondary polycythaemia	3,844

Haematological immunological conditions	Immunodeficiencies	3,800
Haematological immunological conditions	Aplastic anaemias	3,299
Haematological immunological conditions	Sarcoidosis	2,633
Haematological immunological conditions	Other haemolytic anaemias	629
Haematological immunological conditions	Other anaemias	252
Haematological immunological conditions	Iron deficiency anaemia	96
Haematological immunological conditions	Vitamin B12 deficiency anaemia	<5
Infectious diseases	Other or unspecified infectious organisms	440,614
Infectious diseases	HIV	421,121
Infectious diseases	Bacterial diseases excluding TB	367,531
Infectious diseases	Urinary tract infections	299,209
Infectious diseases	Viral diseases excluding chronic hepatitis hiv	183,208
Infectious diseases	Infections of other or unspecified organs	101,009
Infectious diseases	Lower respiratory tract infections	70,990
Infectious diseases	Infection of other or unspecified genitourinary system	37,861
Infectious diseases	Infection of skin and subcutaneous tissues	34,607
Infectious diseases	Infections of the digestive system	27,563
Infectious diseases	Eye infections	17,540
Infectious diseases	Other nervous system infections	14,355
Infectious diseases	Infection of male genital system	13,697
Infectious diseases	Infections of the heart	6,659
Infectious diseases	Infection of anal and rectal regions	6,630
Infectious diseases	Infection of bones and joints	6,425
Infectious diseases	Infection of liver	5,151
Infectious diseases	Meningitis	3,917
Infectious diseases	Chronic viral hepatitis	3,893
Infectious diseases	Septicaemia	3,631
Infectious diseases	Ear and upper respiratory tract infections	3,389
Infectious diseases	Parasitic infections	2,997
Infectious diseases	Mycoses	2,927
Infectious diseases	Tuberculosis	1,796
Infectious diseases	Encephalitis	888
Infectious diseases	Rheumatic fever	363
Mental health disorders	Depression	1,151,654
Mental health disorders	Bipolar affective disorder and mania	710,688
Mental health disorders	Anxiety disorders	479,657
Mental health disorders	Other psychoactive substance misuse	228,296
Mental health disorders	Alcohol problems	157,548
Mental health disorders	Dementia	101,527
Mental health disorders	Intellectual disability	79,640
Mental health disorders	Autism and aspergers syndrome	60,595
Mental health disorders	Hyperkinetic disorders	41,197
Mental health disorders	Schizophrenia schizotypal and delusional disorders	33,377
Mental health disorders	Anorexia and bulimia nervosa	5,812
Mental health disorders	Personality disorders	279
Mental health disorders	Delirium not induced by alcohol and other psychoactive substances	86
Musculoskeletal conditions	Osteoporosis	99,307
Musculoskeletal conditions	Spondylosis	94,560
Musculoskeletal conditions	Fracture of wrist	86,411
Musculoskeletal conditions	Carpal tunnel syndrome	81,606
Musculoskeletal conditions	Enthesopathies synovial disorders	71,593
Musculoskeletal conditions	Fracture of hip	45,259
Musculoskeletal conditions	Rheumatoid arthritis	42,069
Musculoskeletal conditions	Spinal stenosis	41,694
Musculoskeletal conditions	Intervertebral disc disorders	29,904
Musculoskeletal conditions	Polymyalgia rheumatica	16,747
Musculoskeletal conditions	Fibromatoses	14,523
Musculoskeletal conditions	Spondylolisthesis	13,928
Musculoskeletal conditions	Collapsed vertebra	12,101
Musculoskeletal conditions	Psoriatic arthropathy	8,309
Musculoskeletal conditions	Giant cell arteritis	4,170

Musculoskeletal conditions	Sjogrens disease	4,154
Musculoskeletal conditions	Enteropathic arthropathy	829
Musculoskeletal conditions	Gout	67
Musculoskeletal conditions	Systemic sclerosis	21
Musculoskeletal conditions	Lupus erythematosus local and systemic	<5
Neurological conditions	Epilepsy	92,492
Neurological conditions	Postviral fatigue syndrome neurasthenia and fibromyalgia	54,819
Neurological conditions	Peripheral neuropathies excluding cranial nerve and carpal tunnel syndromes	41,829
Neurological conditions	Diabetic neurological complications	32,170
Neurological conditions	Parkinson's disease	15,819
Neurological conditions	Bell's palsy	13,461
Neurological conditions	Intracranial hypertension	5,993
Neurological conditions	Disorders of autonomic nervous system	5,228
Neurological conditions	Trigeminal neuralgia	4,860
Neurological conditions	Essential tremor	3,876
Neurological conditions	Myasthenia gravis	2,415
Neurological conditions	Motor neuron disease	1,299
Perinatal conditions	Slow foetal growth or low birth weight	88,030
Perinatal conditions	Prematurity	53,323
Perinatal conditions	Congenital malformations of cardiac septa	32,115
Perinatal conditions	High birth weight	26,102
Perinatal conditions	Post-term infant	14,760
Perinatal conditions	Patent ductus arteriosus	9,274
Perinatal conditions	Downs syndrome	4,341
Perinatal conditions	Intrauterine hypoxia	3,992
Perinatal conditions	Spina bifida	1,777
Skin conditions	Dermatitis atopc contact other unspecified	154,275
Skin conditions	Pilonidal cyst sinus	30,120
Skin conditions	Actinic keratosis	15,970
Skin conditions	Psoriasis	7,992
Skin conditions	Hidradenitis suppurativa	5,717
Skin conditions	Acne	3,696
Skin conditions	Rosacea	2,343
Skin conditions	Lichen planus	108
Skin conditions	Seborrheic dermatitis	<5

Supplementary table 4: Demographic overview of deceased patients

Table is stratified by individuals identified with or without a formal COVID-19 diagnosis listed on the death certificate, and/or COVID-19 inpatient deaths, as compared with the total population of all patients with a COVID-19 event.

	Fatal with COVID-19 as cause	Fatal without COVID-19 as cause	COVID-19 Inpatient death	COVID-19 Death no hospital contact	COVID-19 Death no hospital contact - Wave 1	COVID-19 Death no hospital contact - Wave 2	All COVID events
n	139818 (1.9)	15486 (0.2)	99938 (1.4)	43814 (0.6)	16203 (6.4)	19677 (0.7)	7244925 (100)
Sex							
Female	63247 (45.2)	7456 (48.1)	40913 (40.9)	23970 (54.7)	8772 (54.1)	11370 (57.8)	3877807 (53.5)
Unknown	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Age							
Under 18	44 (0)	26 (0.2)	49 (0)	27 (0.1)	6 (0)	12 (0.1)	1456663 (20.1)
Age 18 - 29	237 (0.2)	52 (0.3)	197 (0.2)	101 (0.2)	25 (0.2)	36 (0.2)	1458665 (20.1)
Age 30 - 49	3060 (2.2)	466 (3)	2426 (2.4)	987 (2.3)	215 (1.3)	433 (2.2)	2222207 (30.7)
Age 50 - 69	21830 (15.6)	2444 (15.8)	18436 (18.4)	4749 (10.8)	1274 (7.9)	2038 (10.4)	1485351 (20.5)
>= 70	114647 (82)	12498 (80.7)	78830 (78.9)	37950 (86.6)	14683 (90.6)	17158 (87.2)	622039 (8.6)
Unknown	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Ethnicity							
White	121864 (87.2)	14190 (91.6)	85533 (85.6)	39909 (91.1)	14856 (91.7)	17931 (91.1)	5898279 (81.4)
Asian or asian british	9726 (7)	684 (4.4)	8071 (8.1)	1794 (4.1)	500 (3.1)	879 (4.5)	714168 (9.9)
Black or black british	4372 (3.1)	299 (1.9)	3446 (3.4)	1018 (2.3)	402 (2.5)	424 (2.2)	241053 (3.3)
Chinese	342 (0.2)	30 (0.2)	262 (0.3)	100 (0.2)	42 (0.3)	37 (0.2)	21758 (0.3)
Mixed and others	2507 (1.8)	198 (1.3)	1995 (2)	577 (1.3)	250 (1.5)	219 (1.1)	279733 (3.9)
Unknown ethnicity	1007 (0.7)	85 (0.5)	631 (0.6)	416 (0.9)	153 (0.9)	187 (1)	89934 (1.2)
IMD Fifths (%)							
1 (most deprived)	33397 (23.9)	3574 (23.1)	25202 (25.2)	8909 (20.3)	3359 (20.7)	3802 (19.3)	1607009 (22.2)
5 (least deprived)	23024 (16.5)	2713 (17.5)	15533 (15.5)	8333 (19)	3180 (19.6)	3874 (19.7)	1334226 (18.4)
Unknown	120 (0.1)	15 (0.1)	62 (0.1)	57 (0.1)	25 (0.2)	23 (0.1)	5272 (0.1)
COVID-19 events							
COVID-19 positive test	120326 (86.1)	6580 (42.5)	90154 (90.2)	26486 (60.5)	6321 (39)	15730 (79.9)	6778342 (93.6)
GP COVID-19 diagnosis	55176 (39.5)	6059 (39.1)	33564 (33.6)	20903 (47.7)	5899 (36.4)	11247 (57.2)	3056132 (42.2)
COVID-19 admission	102913 (73.6)	8579 (55.4)	99920 (100)	0 (0)	0 (0)	0 (0)	460737 (6.4)
ICU admission	17833 (12.8)	722 (4.7)	18465 (18.5)	0 (0)	0 (0)	0 (0)	48847 (0.7)
NIV treatment	26895 (19.2)	895 (5.8)	27442 (27.5)	0 (0)	0 (0)	0 (0)	69090 (1)
IMV treatment	12879 (9.2)	490 (3.2)	13459 (13.5)	0 (0)	0 (0)	0 (0)	25928 (0.4)
ECMO treatment	246 (0.2)	1 (0)	254 (0.3)	0 (0)	0 (0)	0 (0)	696 (0)
Fatal with COVID-19 as cause	139818 (100)	0 (0)	91164 (91.2)	36904 (84.2)	15000 (92.6)	17128 (87)	139818 (1.9)
Fatal without COVID-19 as cause	0 (0)	15486 (100)	6058 (6.1)	6907 (15.8)	1201 (7.4)	2549 (13)	15486 (0.2)
COVID-19 inpatient death	91164 (65.2)	6058 (39.1)	99938 (100)	17 (0)	15 (0.1)	0 (0)	99938 (1.4)

Supplementary table 5: Primary diagnosis on death certificate for deceased patients

Table shows the top ten most frequent primary diagnosis on the death certificate after removal of duplicates (defined as entries with the same ID, date and underlying cause of death) and registrations with a null underlying cause of death for the 139,818 individuals with COVID-19 on the death certificate, and 15,486 dying without COVID-19 on the death certificate within 28 days of a COVID-19 event.

COVID on the death certificate			Without COVID on the death certificate		
ICD10	**Description**	**N (%)**	**ICD10**	**Description**	**N (%)**
U071	COVID-19, virus identified	121897 (87.2%)	F03	Unspecified dementia	1008 (6.5%)
U072	COVID-19, virus not identified	3529 (2.5%)	C349	Cancer of bronchus and lung	809 (5.2%)
F03	Unspecified dementia	1336 (1%)	J189	Pneumonia	794 (5.1%)
I259	Chronic ischemic heart disease	832 (0.6%)	I259	Chronic ischemic heart disease	525 (3.4%)
C349	Cancer of bronchus and lung	786 (0.6%)	J440	Chronic obstructive pulmonary disease	495 (3.2%)
I64	Stroke	732 (0.5%)	I64	Stroke	490 (3.2%)
G309	Alzheimer's disease	710 (0.5%)	G309	Alzheimer's disease	444 (2.9%)
I219	Acute myocardial infarction	634 (0.5%)	I219	Acute myocardial infarction	431 (2.8%)
F019	Dementia	552 (0.4%)	C61	Malignant neoplasm of prostate	348 (2.2%)
W19	Unspecified fall	411 (0.3%)	F019	Dementia	341 (2.2%)

Supplementary Table 6: Primary diagnoses of COVID-19 hospitalisations identified from HES APC

Table shows the top 10 most common primary diagnoses in hospitalisations identified from HES APC with COVID-19 at any diagnostic position. Primary diagnosis is defined as the diagnosis appearing in the primary position for the first episode in a patient’s hospital admission. A total of 3,726 unique ICD-10 codes were identified in primary position in the first episode of a COVID-19 admission.

ICD-10 Code	Description	Episodes	Individuals	% individuals
U071	COVID-19, virus identified	252,599	252,088	56.6
U072	COVID-19, virus not identified	20,572	20,530	4.6
N390	Urinary tract infection, site not specified	5,578	5,566	1.3
A419	Sepsis, unspecified	4,953	4,944	1.1
J181	Lobar pneumonia, unspecified	4,633	4,627	1
R296	Tendency to fall, not elsewhere classified	4,491	4,481	1
S720	Fracture of neck of femur	3,750	3,735	0.8
N179	Acute renal failure, unspecified	3,556	3,542	0.8
R69X	Unknown and unspecified causes of morbidity	3,190	3,178	0.7
I500	Congestive heart failure	3,114	3,108	0.7

*U072 is intended for use when COVID-19 is diagnosed clinically or epidemiologically but laboratory testing is inconclusive or not available

RECORD Statement

The Reporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Title & Abstract/Design Abstract/Participants Abstract/Design
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Introduction		
Objectives	3	State specific objectives, including any prespecified hypotheses	Introduction/final paragraph		
Methods					
Study Design	4	Present key elements of study design early in the paper	Method/Study design and EHR data sources		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods/Design, Population		

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Methods, Figure 1 & 2, Supplementary Table 1, Github Comorbidities: Wood et al. 2021, Kuan et al. 2019 Cross-EHR source concordance and consistency with established knowledge discussed Flow diagram included of datasources, linkage provided by NHS-D and referenced, therefore not explicitly described
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Methods, Supplementary Table 1, Github
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Methods, Supplementary Table 1		

			Cross-EHR source concordance shown in Figure 3 and Supplementary Figure 2. Temporal coherence in Figure 4 and Supplementary Figure 1		
Bias	9	Describe any efforts to address potential sources of bias	Discussion		
Study size	10	Explain how the study size was arrived at	Methods/Population, Figure 1		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Methods/Covariates & comorbidities *Added discretisation of Age*		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed	Methods/Statistical analyses. No adjustment for confounding, discussed in discussion. Number of missing data fields reported for all key variables Follow up of 28 days for		

		Cross-sectional study - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	pandemic wave analysis described in method NA NA NA		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Methods/Ethical & regulatory approvals (Access) Cleaning -> Github full analysis code
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Methods/Design
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	Methods/Population, Figure 1	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical,	Results Table 1		

		social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	Results Table 1 Results		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	Results Table 1		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Results; No adjustment for confounding, discussed in discussion		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Trajectory analysis reported in main results		
Discussion					
Key results	18	Summarise key results with reference to study objectives	Discussion		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion/Strengths and limitations

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion/Comparison with previous findings, Strengths and limitations		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion/Strengths and limitations		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Funding		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Full analysis code available on GitHub. Protocols available via CVD-COVID-UK consortium home page.

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

Supplementary references

1 General Practice Extraction Service (GPES) Data for pandemic planning and research: a guide for analysts and users of the data. NHS Digital. https://digital.nhs.uk/coronavirus/gpes-data-for-pandemic-planning-and-research/guide-for-analysts-and-users-of-the-data (accessed Jan 20, 2022).

2 Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69**: 4–14.

3 Department of Health and Social Care. Payment by Results in the NHS: a simple guide. GOV.UK. 2013; published online March 25. https://www.gov.uk/government/publications/simple-guide-to-payment-by-results (accessed Jan 25, 2022).

4 Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L. Understanding Hospital episode statistics (HES). *London, UK: CLOSER* 2017. https://www.closer.ac.uk/wp-content/uploads/CLOSER-resource-understanding-hospital-episode-statistics-2018.pdf.

5 Burns EM, Rigby E, Mamidanna R, *et al.* Systematic review of discharge coding accuracy. *J Public Health* 2012; **34**: 138–48.

6 The processing cycle and HES data quality. NHS Digital. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/the-processing-cycle-and-hes-data-quality (accessed Jan 25, 2022).

7 Campbell A. Quality of mortality data during the coronavirus pandemic, England and Wales - Office for National Statistics. 2020; published online Dec 3. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/qualityofmortalitydataduringthecoronaviruspandemicenglandandwales/2020 (accessed Jan 26, 2022).

8 NHS Data Quality Maturity Index. NHS Digital. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/data-quality (accessed March 13, 2019).

9 Wood A, Denholm R, Hollings S, *et al.* Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource. *BMJ* 2021; **373**: n826.

10 [MI] National Data Opt-out, September 2021. NHS Digital. https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/september-2021 (accessed Jan 20, 2022).

11 Baker C. Population estimates & GP registers: why the difference? 2016; published online Dec 12. https://commonslibrary.parliament.uk/population-estimates-gp-registers-why-the-difference/ (accessed Jan 20, 2022).

12 Summary of latest statistics. GOV.UK.

https://www.gov.uk/government/statistics/immigration-statistics-year-ending-september-2021/summary-of-latest-statistics (accessed Jan 20, 2022).

13 Clare T, Twohig KA, O'Connell A-M, Dabrera G. Timeliness and completeness of laboratory-based surveillance of COVID-19 cases in England. *Public Health* 2021; **194**: 163–6.

14 Denaxas S, Gonzalez-Izquierdo A, Direk K, *et al.* UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc* 2019; **26**: 1545–59.

15 Data Access Request Service (DARS). https://digital.nhs.uk/services/data-access-request-service-dars (accessed May 21, 2021).