Identifying long COVID using electronic health records: a national observational cohort study in Scotland: Supplementary Material

Contents

Tables

Figures

Reporting guidelines

The manuscript was guided by the Strengthening the Reporting of Observational Studies in Epidemiology checklist (Table S1).

Table S1: STROBE Statement—Checklist of items that should be included in reports of cohort studies [Error! Bookmark not defined.**]**

	Item.	Recommendation	Page No.
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Title page
		(*b*) Provide in the abstract an informative and balanced summary	Page 3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 4
Methods			
Study design	4	Present key elements of study design early in the paper	Page 4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	Page 6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 5-7
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 more than one group	Page 5-7
Bias	9	Describe any efforts to address potential sources of bias	Page 7
Study size	10	Explain how the study size was arrived at	Page 5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Page 7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 6-8
		(*b*) Describe any methods used to examine subgroups and interactions	Page 8
		(*c*) Explain how missing data were addressed	Page 7
		(*d*) If applicable, explain how loss to follow-up was addressed	-
		(*e*) Describe any sensitivity analyses	Page 8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—e.g. numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Table 1
		(b) Give reasons for non-participation at each stage	Page 10
		(c) Consider use of a flow diagram	Figure S1
Descriptive data	14*	(a) Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders	Table 1-2
		(b) Indicate number of participants with missing data for each variable of interest	Table 1
		(c) Summarise follow-up time (e.g., average and total amount)	Page 8
Outcome data	15*	Report numbers of outcome events or summary measures over time	Figure 1
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 1-2
		(*b*) Report category boundaries when continuous variables were categorized	Table 1
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—analyses of subgroups and interactions, and sensitivity analyses	Page 8
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 8-9
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	Page 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Page 12
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 10-12
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	Page 13

Linked datasets

Pseudonymised identifiers of National Health Service (NHS) Scotland's Community Healthcare Index (CHI) were used to link the following datasets:

- Primary care
  - Primary care health records and demographic data
  - Out of hours data (Out of Hours Data Mart)

- Secondary care
  - In-patient data from Scottish Morbidity Record 01 (SMR-01)
  - Out-patient data from Scottish Morbidity Record 00 (SMR-00)
  - Intensive care admissions from the Scottish Intensive Care Society Audit Group (SICSAG)
  - Unscheduled visits to Accident & Emergency (A&E)
  - NHS 24 calls

- Prescribing
  - Dispensed prescriptions from the Prescribing Information System (PIS)

- Testing and vaccinations
  - Reverse transcriptase polymerase chain reaction (RT-PCR) and lateral flow testing (LFT) data from the Electronic Communication or Surveillance in Scotland (ECOSS)
  - Whole Genome Sequencing data from the Centre of Genomics (COG)
  - Records of vaccinations, shielding, and immunocompromised individuals extracted from Public Health Scotland's (PHS) Turas Vaccination Management Tool (TVMT

- Deaths
  - Death registry data from the National Records of Scotland (NRS)

Code

All code underpinning the analyses presented in this study is publicly available at https://github.com/EAVE-II/Long-COVID.

Table S2: Code lists of grouped clinical codes (Read version 2)

Clinical code group	Read code	Description
Abdominal pain	1969.	Abdominal pain
	R090.	[D]Abdominal pain
	R090y	[D]Other specified abdominal pain
	R090z	[D]Abdominal pain NOS
	Ryu11	[X]Other and unspecified abdominal pain
Anorexia	161..	Appetite symptom
	1612.	Appetite loss - anorexia
	1615.	Reduced appetite
	161Z.	Appetite symptom NOS
	R030.	[D]Anorexia
	R0300	[D]Appetite loss
	1611.	Appetite normal
	1613.	Appetite increased
Anxiety	1B1..	General nervous symptoms
	1B13.	Anxiousness
	8G94.	Anxiety management training
	E200.	Anxiety states
	E2000	Anxiety state unspecified
	E2001	Panic disorder
	E2002	Generalised anxiety disorder
	E2003	Anxiety with depression
	E2004	Chronic anxiety
	E2005	Recurrent anxiety
	E200z	Anxiety state NOS
	E201A	Dissociative reaction unspecified
	E205.	Neurasthenia - nervous debility
	E207.	Hypochondriasis
	E20y.	Other neurotic disorders
	E26..	Physiological malfunction arising from mental factors
	E2613	Psychogenic hyperventilation
	E28..	Acute reaction to stress
	E2830	Acute situational disturbance
	E29..	Adjustment reaction
	E292.	Adjustment reaction, predominant disturbance other emotions
	E2924	Adjustment reaction with anxious mood
	E29y1	Other post-traumatic stress disorder
	E29z.	Adjustment reaction NOS
	Eu40z	[X]Phobic anxiety disorder, unspecified
	Eu41.	[X]Other anxiety disorders
	Eu410	[X]Panic disorder [episodic paroxysmal anxiety]
	Eu411	[X]Generalized anxiety disorder
	Eu412	[X]Mixed anxiety and depressive disorder
	Eu413	[X]Other mixed anxiety disorders
	Eu41y	[X]Other specified anxiety disorders
	Eu41z	[X]Anxiety disorder, unspecified
	Eu420	[X]Predominantly obsessional thoughts or ruminations

Clinical code group	Read code	Description
	Eu43.	[X]Reaction to severe stress, and adjustment disorders
	Eu430	[X]Acute stress reaction
	Eu431	[X]Post - traumatic stress disorder
	Eu432	[X]Adjustment disorders
	Eu446	[X]Dissociative anaesthesia and sensory loss
	Eu44z	[X]Dissociative [conversion] disorder, unspecified
	Eu45.	[X]Somatoform disorders
	Eu452	[X]Hypochondriacal disorder
	Eu453	[X]Somatoform autonomic dysfunction
	Eu45y	[X]Other somatoform disorders
	Eu45z	[X]Somatoform disorder, unspecified
	Eu460	[X]Neurasthenia
	Eu46y	[X]Other specified neurotic disorders
	Eu46z	[X]Neurotic disorder, unspecified
	Eu42.	[X]Obsessive - compulsive disorder
	E2900	Grief reaction
	E21..	Personality disorders
	E203.	Anancastic neurosis
Back and Neck Pain	16C..	Backache symptom
	16C2.	Backache
	16C5.	C/O - low back pain
	16C6.	Back pain without radiation NOS
	16C7.	C/O - upper back ache
	16C9.	Chronic low back pain
	16CA.	Mechanical low back pain
	16CZ.	Backache symptom NOS
	1A53.	Lumbar ache - renal
	N12..	Acute back pain - disc
	N131.	Cervicalgia - pain in neck
	N141.	Pain in thoracic spine
	N142.	Acute back pain - lumbar
	N143.	Acute back pain with sciatica
	N145.	Backache, unspecified
	N14y.	Other back symptoms
Blood Tests Biochemistry	44D6.	Liver function test
	44I8.	Serum calcium
	44I9.	Serum inorganic phosphate
Blood Tests Endocrine	442..	Thyroid hormone tests
	44AJ.	Plasma parathyroid hormone level
Blood Tests Glucose	42W5.	Haemoglobin A1c level - IFCC standardised
	44g..	Plasma glucose level
Blood Tests Haematology	424..	Full blood count - FBC
	42R4.	Serum ferritin
	42R7.	Serum iron level
	42YD.	B12/folate level
	42Z..	Haematology NOS
Blood Tests Inflammation	42B6.	Erythrocyte sedimentation rate

Clinical code group	Read code	Description
	44CC.	Plasma C reactive protein
	44CC0	C reactive protein normal
	44CC1	C reactive protein abnormal
	44CS.	Serum C reactive protein level
Blood Tests Renal	44JB.	Urea and electrolytes
	451E.	GFR calculated abbreviated MDRD
Breathless	17...	Respiratory symptoms
	173..	Breathlessness
	1732.	Breathless - moderate exertion
	1733.	Breathless - mild exertion
	1734.	Breathless - at rest
	1735.	Orthopnoea symptom
	1736.	Paroxysmal nocturnal dyspnoea
	1737.	Wheezing symptom
	1738.	Difficulty breathing
	1739.	Shortness of breath
	173C.	Short of breath on exertion
	173D.	Nocturnal dyspnoea
	173F.	Short of breath dressing/undressing
	173G.	Breathless - strenuous exertion
	173H.	MRC Breathlessness Scale: grade 1
	173I.	MRC Breathlessness Scale: grade 2
	173J.	MRC Breathlessness Scale: grade 3
	173K.	MRC Breathlessness Scale: grade 4
	173L.	MRC Breathlessness Scale: grade 5
	173Z.	Breathlessness NOS
	1J70.	Suspected asthma
	2322.	O/E - dyspnoea
	2323.	O/E - orthopnoea
	2324.	O/E - respiratory distress
	663q.	Asthma daytime symptoms
	8Hlj.	Referr to British Lung Foundation breathe easy support group
	H5853	Adult respiratory distress syndrome
	Q30..	Respiratory distress syndrome
	R0602	[D]Orthopnoea
	R0606	[D]Respiratory distress
	R0608	[D]Shortness of breath
	R060A	[D]Dyspnoea
	R060D	[D]Breathlessness
Chest X-ray	535..	Standard chest X-ray
	5351.	Standard chest X-ray requested
	5352.	Standard chest xray normal
	5353.	Standard chest xray abnormal
	535Z.	Standard chest xray NOS
	536..	Soft tissue X-ray chest
	5361.	Soft tissue X-ray chest normal
	5362.	Soft tissue X-ray chest abnormal

Clinical code group	Read code	Description
	5364.	Soft tiss.X-ray lung/bronchus
	536Z.	Soft tissue X-ray chest NOS
	5637.	Tomography – lungs
	68C1.	Screening chest X-ray
Chest Pain	182..	Chest pain
	1822.	Central chest pain
	1825.	Pleuritic pain
	1827.	Pleurodynia
	1828.	Atypical chest pain
	1829.	Retrosternal pain
	182A.	Chest pain on exertion
	182Z.	Chest pain NOS
	N2410	Intercostal myalgia
	N33zE	Costochondritis
	N33zz	Costochondritis NOS
	R065.	[D]Chest pain
	R0650	[D]Chest pain, unspecified
	R0652	[D]Anterior chest wall pain
	R0653	[D]Painful respiration NOS
	R0654	[D]Pleuritic pain
	R0656	[D]Chest discomfort
	R0657	[D]Chest pressure
	R0658	[D]Chest tightness
	R065A	[D]Musculoskeletal chest pain
	R065B	[D]Non cardiac chest pain
	R065C	[D]Retrosternal chest pain
	R065D	[D]Central chest pain
	R065z	[D]Chest pain NOS
	Ryu04	[X]Other chest pain
Cognitive Impairment	28E..	Cognitive decline
	28E0.	Mild cognitive impairment
	28E1.	Moderate cognitive impairment
	28E2.	Severe cognitive impairment
	28E3.	Cognitive impairment
Confusion	1B1A.	Amnesia symptom
	1B1A1	Short-term memory loss
	1BR..	Reduced concentration
	1BR0.	Reduced concentration span
	1BW..	Poor concentration
	1S21.	Disturbance of memory for order of events
	1S23.	Memory impairment
	2232.	O/E - confused
	2841.	Confused
	E030.	Acute confusional state
	E0301	Acute confusional state, of infective origin
	E030z	Acute confusional state NOS
	E031.	Subacute confusional state

Clinical code group	Read code	Description
	E042.	Chronic confusional state
	E132.	Reactive confusion
	E2A10	Mild memory disturbance
	Eu...	[X]Mental and behavioural disorders
	Eu04.	[X]Delirium, not induced by alcohol and other psychoactive subs
	Eu04y	[X]Other delirium
	Eu057	[X]Mild cognitive disorder
	G655.	Transient global amnesia
	R009.	[D]Confusion
	R0090	[D]Toxic confusional state
	R00z0	[D]Amnesia (retrograde)
	R00zD	[D]Restlessness and agitation
	R00zX	[D]Disorientation, unspecified
	Ryu50	[X]Other amnesia
	Ryu55	[X]Other symptoms and signs involving emotional state
	ZV40.	[V]Behavioural problems
Consult Mental Health	1S...	Mental and psychological observations
	1S4..	Mood observations
	38C1.	Mental health assessment
	665..	Psych. disorder monitoring
	67H..	Lifestyle counselling
	6896.	Depression screening using questions
	6A6..	Mental health review
Sick Note	9C8..	(Sickness notif GP) or (LOC1/2/3)
	9C82.	LOC 2-sickness notification
	9C83.	LOC 3-sickness payment record
	9C8Z.	LOC 1/2/3 - NOS
	9D1..	MED3 - doctor's statement
	9D11.	MED3 issued to patient
	9D12.	MED3 duplicate issued
	9D15.	eMED3 (2010) new statement issued, not fit for work
	9D16.	eMED3 (2010) new statement issued, may be fit for work
	9D17.	eMED3 (2010) duplicate issued, not fit for work
	9D19.	MED3 (2010) issued by hand, not fit for work
	9D1Z.	MED3 - NOS
	9D5..	Private sickness certificate
	9D7..	Forces sickness on leave certificate
	9D71.	Forces sickness on leave certification - fee not paid
	9D72.	Forces sickness on leave certification - receipt given
	9D7Z.	Forces sick on leave cert NOS
	Ua0Ur	On sick leave from work
	XaBHC	LOC1/2/3 - notification of sickness status
	XaBHd	MED3 status
	XaBHi	Forces sickness on leave certification status
	XaBIc	Sickness certificates
	XaBQ8	Sickness payment claim status
	XaBu9	Sickness notification-of GP

Clinical code group	Read code	Description
	XaCF3	SC1 - self certificate
	XaCFk	LOC 1/2/3 status
	XaK2J	Sick note generated from secondary care done by practice
	XaX1E	eMED3 (2010) new statement issued, not fit for work
	XaX1K	eMED3 (2010) new statement issued, may be fit for work
	XaX1L	eMED3 (2010) duplicate issued, not fit for work
	XaX1M	eMED3 (2010) duplicate issued, may be fit for work
	XaX1R	MED3 (2010) issued by hand, not fit for work
	XaX1S	MED3 (2010) issued by hand, may be fit for work
	XaXf1	MED3 (2010) issued to patient
	XaXf2	MED3 (2010) duplicate issued
	XaXf3	MED3 (2010) issued - recommend phased return to work
	XaXf4	MED3 (2010) issued - recommend altered hours
	XaXf5	MED3 (2010) issued - recommend amended duties
	XaXf7	MED3 (2010) issued - recommend workplace adaptation
	XE2by	LOC1/2/3- notific. of sickness
	Y0898	Recorded cause of certified sickness absence from work
	Y08c1	Duration of sickness certificate (Weeks)
	Y08c2	Duration of sickness certificate (Months)
	Y0d09	MED3 - Specified conditions
	Y1712	Duration of sickness certificate (Days)
	Y2854	Employment status - Long term sickness
	Y6858	Forces sickness on leave cert.
	ZV680	[V]Issue of medical certificate
	9D18.	eMED3 (2010) duplicate issued, may be fit for work
Cough	171..	Cough
	1712.	Dry cough
	1713.	Productive cough -clear sputum
	1714.	Productive cough -green sputum
	1715.	Productive cough-yellow sputum
	1716.	Productive cough NOS
	1717.	Night cough present
	1719.	Chesty cough
	171A.	Chronic cough
	171B.	Persistent cough
	171E.	Unexplained cough
	171F.	Cough with fever
	171K.	Barking cough
	171L.	Cough on exercise
	171Z.	Cough symptom NOS
	173B.	Nocturnal cough / wheeze
	R062.	[D]Cough
	R0621	[D]Episodic dry cough
COVID	1JX1.	Suspected disease caused by 2019-nCoV
	4J3R1	2019-nCoV (novel coronavirus) detected
	8HkjG	Signposting to Your COVID Recovery
	99116	Suspected case of the disease COVID-19

Clinical code group	Read code	Description
	A795.	Coronavirus infection
	A7951	Disease caused by 2019-nCoV
	AyuDC	Coronavirus infection, unspecified
	CO303	Confirmed 2019-nCoV
COVID Possible	1JX..	Suspected coronavirus infection
	65PW1	Exposure to 2019-nCoV
	8CAO.	Advice given about 2019-nCoV (novel coronavirus) infection
Depression	1B17.	Depressed
	1B1U.	Symptoms of depression
	1BT..	Depressed mood
	212S.	Depression resolved
	8HHq.	Referral for guided self-help for depression
	9H90.	Depression annual review
	9H91.	Depression medication review
	9H92.	Depression interim review
	E112.	Agitated depression
	E1122	Single major depressive episode, moderate
	E1123	Single major depressive episode, severe, without psychosis
	E112z	Single major depressive episode NOS
	E113.	Endogenous depression - recurrent
	E1130	Recurrent major depressive episodes, unspecified
	E1132	Recurrent major depressive episodes, moderate
	E1137	Recurrent depression
	E118.	Seasonal affective disorder
	E135.	Agitated depression
	E204.	Neurotic depression reactive type
	E2112	Depressive personality disorder
	E290z	Brief depressive reaction NOS
	E2B..	Depressive disorder NEC
	E2B1.	Chronic depression
	Eu32.	[X]Depressive episode
	Eu320	[X]Mild depressive episode
	Eu321	[X]Moderate depressive episode
	Eu322	[X]Severe depressive episode without psychotic symptoms
	Eu324	[X]Mild depression
	Eu32y	[X]Other depressive episodes
	Eu32z	[X]Depressive episode, unspecified
	Eu33.	[X]Recurrent depressive disorder
	Eu331	[X]Recurrent depressive disorder, current episode moderate
	Eu332	[X]Recurr depress disorder cur epi severe without psyc sympt
	Eu334	[X]Recurrent depressive disorder, currently in remission
	Eu33y	[X]Other recurrent depressive disorders
	Eu33z	[X]Recurrent depressive disorder, unspecified
	Eu341	[X]Dysthymia
	Eu3y1	[X]Other recurrent mood affective disorders
	Eu530	[X]Mild mental/behav disorder assoc with the puerperium NEC
Diarrhoea	19...	Gastrointestinal symptoms

Clinical code group	Read code	Description
	19F..	Diarrhoea symptoms
	19F2.	Diarrhoea
	19FZ.	Diarrhoea & vomiting, symptom
	19G..	Diarrhoea and vomiting
	A083.	Diarrhoea of presumed infectious origin
	J4zz.	Diarrhoea - presumed non-infectious
	19F1.	Diarrhoea not present
Dizziness	1B5..	Dizziness symptom
	1B53.	Dizziness present
	1B55.	Dizziness on standing up
	1B56.	Vertigo
	1B6..	Faint symptom
	1B62.	Syncope/vasovagal faint
	1B68.	Felt faint
	A78y0	Epidemic vertigo
	F561.	Other and unspecified peripheral vertigo
	F5610	Unspecified peripheral vertigo
	F5611	Benign paroxysmal positional vertigo or nystagmus
	F5614	Aural vertigo
	F5615	Benign paroxysmal positional vertigo
	F561z	Other peripheral vertigo NOS
	F562.	Vertigo of central origin
	FyuQ1	[X]Other peripheral vertigo
	R0021	[D]Fainting
	R004.	[D]Dizziness and giddiness
	R0040	[D]Dizziness
	R0042	[D]Light-headedness
	R0043	[D]Vertigo NOS
	R0044	[D]Acute vertigo
	R004z	[D]Dizziness and giddiness NOS
Dyspepsia	195..	Indigestion symptoms
	1953.	Waterbrash
	1954.	Indigestion
	1955.	Heartburn symptom
	1957.	Gastric reflux
	1958.	Undiagnosed dyspepsia
	195Z.	Indigestion symptom NOS
	J10y4	Gastro-oesophageal reflux
	J16y4	Dyspepsia
	R071.	[D]Heartburn
	R0710	[D]Pyrosis
	R0711	[D]Waterbrash
	R071z	[D]Heartburn NOS
Earache	1C3..	Earache symptoms
	1C32.	Unilateral earache
	1C33.	Bilateral earache
	1C34.	Irritation of ear

Clinical code group	Read code	Description
	1C3Z.	Earache symptom NOS
	F502z	Otitis externa NOS
	F587.	Otalgia
	F5872	Referred ear pain
Echocardiogram	32...	Electrocardiography
	321..	ECG - general
	3211.	ECG requested
	3212.	Standard ECG
	3213.	Exercise ECG
	321B.	12 lead ECG
	321Z.	ECG - general - NOS
	32M..	24 Hour ECG
	32Z..	Electrocardiography NOS
	7P0G.	Diagnostic electrocardiography
	7P0G3	Exercise electrocardiography
	7P0Gy	Other specified diagnostic electrocardiography
	7P0Gz	Diagnostic electrocardiography NOS
	7P0P.	Other diagnostic electrocardiography
	8A52.	ECG monitoring
	R1431	[D]Electrocardiogram (ECG) abnormal
Fatigue	138..	Exercise grading
	1381.	Exercise physically impossible
	168..	Fatigue - symptom
	1682.	Fatigue
	1683.	Tired all the time
	1684.	Malaise/lethargy
	1688.	Exhaustion
	F286.	Chronic fatigue syndrome
	R007.	[D]Malaise and fatigue
	R0070	[D]Malaise
	R0071	[D]Fatigue
	R0072	[D]Asthenia NOS
	R0073	[D]Lethargy
	R0074	[D]Post viral debility
	R0075	[D]Tiredness
	R007z	[D]Postoperative depression
	ZV4K6	[V]Burn-out
Fever	165..	Temperature symptoms
	1652.	Feels hot/feverish
	1653.	Fever with sweating
	1657.	Hot flushes
	165Z.	Temperature symptom NOS
General Symptoms	16...	General symptoms
	1662.	Excessive sweating
	1692.	Swollen glands
	16E..	Feels unwell
	16Z..	General symptom

Clinical code group	Read code	Description
	16ZZ.	General symptom NOS
Headache	1B1G.	C/O - a headache
	1BA2.	Generalised headache
	1BA3.	Unilateral headache
	1BA4.	Bilateral headache
	1BA5.	Frontal headache
	1BA6.	Occipital headache
	1BA7.	Parietal headache
	1BA8.	Temporal headache
	1BA9.	Sinus headache
	1BAZ.	Headache site NOS
	1BB..	Headache character
	1BB1.	Aching headache
	1BB2.	Throbbing headache
	1BB3.	Shooting headache
	1BB4.	Morning headache
	1BB5.	Heavy head
	1BBZ.	Headache character NOS
	8B6N.	Migraine prophylaxis
	E2781	Tension headache
	Eu454	[X]Persistent somatoform pain disorder (P) Psychogenic headache (S)
	F26..	Migraine
	F260.	Classical migraine
	F261.	Common migraine
	F2610	Atypical migraine
	F2611	Sick headache
	F261z	Common migraine NOS
	F262.	Migraine variants
	F2620	Cluster headache
	F2621	Horton's (histamine) neuralgia
	F2623	Basilar migraine
	F2624	Ophthalmic migraine
	F2625	Periodic migrainous neuralgia
	F2626	[X]Tension-type headache
	F2627	Chronic paroxysmal hemicrania
	F262z	Migraine variant NOS
	F26y.	Other forms of migraine
	F26y0	Hemiplegic migraine
	F26y1	Ophthalmoplegic migraine
	F26y2	Status migrainosus
	F26y3	Complicated migraine
	F26yz	Other forms of migraine NOS
	F26z.	Migraine NOS
	Fyu5D	Cervicogeneic headache
	Fyu5E	[X]Chronic headache disorder
	R040.	[D]Headache
	R040z	[D]Pain in head NOS

Clinical code group	Read code	Description
Insomnia	1B1B.	Cannot sleep - insomnia
	1B1B0	Initial insomnia
	1B1B1	Middle insomnia
	1B1B2	Late insomnia
	E2741	Transient insomnia
	E2742	Persistent insomnia
	R005.	[D]Sleep disturbances
	R0050	[D]Sleep disturbance, unspecified
	R0052	[D]Insomnia NOS
	R0054	[D]Hypersomnia NOS
Joint and Bone Pain	182B.	Rib pain
	1D130	C/O - pain in toes
	1D131	C/O - pain in big toe
	1M00.	Pain in elbow
	1M01.	Pain in wrist
	1M02.	Shoulder joint painful on movement
	1M10.	Knee pain
	1M11.	Foot pain
	1M12.	Anterior knee pain
	1M13.	Ankle pain
	2H45.	O/E - joint movement painful
	J0464	Temporomandibular joint-pain-dysfunction syndrome
	N094.	Ache in joint
	N0940	Arthralgia of unspecified site
	N0941	Shoulder joint pain
	N0942	Elbow joint pain
	N0943	Wrist joint pain
	N0944	Hand joint pain
	N0945	Coxalgia
	N0946	Knee joint pain
	N0947	Ankle joint pain
	N0949	Arthralgia of multiple joints
	N094A	Arthralgia of shoulder
	N094B	Arthralgia of sternoclavicular joint
	N094D	Arthralgia of elbow
	N094F	Arthralgia of wrist
	N094J	Arthralgia of DIP joint of finger
	N094K	Arthralgia of hip
	N094M	Arthralgia of knee
	N094V	Arthralgia of IP joint of toe
	N094W	Anterior knee pain
	N094z	Arthralgia NOS
	N0960	Other joint symptoms of unspecified site
	N0961	Other joint symptoms of the shoulder region
	N0965	Hip snapping
	N0966	Other joint symptoms of the lower leg
	N0967	Other joint symptoms of the ankle and foot

Clinical code group	Read code	Description
	N096D	Other symptoms - elbow
	N096K	Other symptoms - hip
	N1472	Coccygodynia
	N2172	Metatarsalgia NOS
	N245.	Pain in limb
	N2450	Hand pain
	N2451	Foot pain
	N2457	Shoulder pain
	N33A.	Bone pain
	N33A0	Bony pelvic pain
	N33A1	Clavicle pain
	N2157	Trochanteric bursitis
Long COVID	^ESCT1348645	Post-COVID-19 syndrome
	^ESCT1348648	Ongoing symptomatic COVID-19
	A7955	Ongoing symptomatic COVID-19
	AyuJC	Post-COVID-19 syndrome
Muscle Pain	1DCC.	Aching muscles
	N20..	Polymyalgia rheumatica
	N239.	Myofascial pain syndrome
	N2413	Viral myalgia
	N2454	Calf pain
	R01..	[D]Nervous and musculoskeletal symptoms
	R01z2	[D]Musculoskeletal pain
Nausea	198..	Nausea
	1981.	No nausea
	1982.	Nausea present
	E2642	Cyclical vomiting - psychogenic
	Eu505	[X]Vomiting associated with other psychological disturbances
	J1620	Cyclical vomiting NOS
	J16y5	Functional vomiting
	R070.	[D]Nausea and vomiting
	R0700	[D]Nausea
	R0701	[D]Vomiting
Pain	1D13.	C/O: a pain
	1DC1.	Burning pain
	1DC2.	Aching pain
	1DC6.	Tightening pain
	1DC8.	Generalised pain [symptom]
	1DCA.	Rest pain
	1DCG.	Cramping pain
	1DCZ.	Pain character NOS
	1M...	Pain
	1M0..	Pain in upper limb
	1M1..	Pain in lower limb
	1M51.	Intermittent pain
	1M52.	Chronic pain
	2I18.	O/E - pain

Clinical code group	Read code	Description
	F369.	Complex regional pain syndrome
	N2452	Pain in leg
	N2453	Pain in arm
	R00z2	[D]General aches and pains
	R00zB	[D]Acute pain
	R00zC	[D]Chronic intractable pain
	R090P	[D]Functional abdominal pain syndrome
	Ryu70	[X]Other chronic pain
Palpitations	14AD.	H/O ventricular fibrillation
	14AQ.	History of supraventricular tachycardia
	181..	Palpitations
	1812.	Palpitations
	1814.	"Fluttering" of heart
	181Z.	Palpitations NOS
	1J62.	Suspected arrhythmia
	F2560	Hypsarrhythmia
	G5674	Wolff-Parkinson-White syndrome
	G57..	Cardiac arrhythmias
	G570.	Paroxysmal supraventricular tachycardia
	G574.	Ventricular fibrillation and flutter
	G5740	Ventricular fibrillation
	G576.	Ectopic beats
	G5760	Extrasystoles
	G577.	Sinus arrhythmia
	G57y.	Other cardiac dysrhythmias
	G57y6	Nodal rhythm disorder
	G57y9	Supraventricular tachycardia NOS
	G57z.	Cardiac dysrhythmia NOS
	R051.	[D]Palpitations
	R051z	[D]Palpitations NOS
Physiotherapy	03J5.	Occupational therapist
	8E...	Physiotherapy/remedial therapy
	8F3..	Occupational therapy
	8H77.	Refer to physiotherapist
	9NJk.	In-house physiotherapy first appointment
Rash	1D14.	C/O: a rash
	1N...	Symptoms of skin and integumentary tissue
	2114.	O/E - a rash
	2227.	O/E - rash present
	2F0..	O/E - discoid rash
	M28z.	Urticaria NOS
	M2y42	Vesicular eruption
	R021.	[D]Rash and other nonspecific skin eruption
	R021z	[D]Rash and other nonspecific skin eruption NOS
Respiratory Infection	H05z.	Upper respiratory infection NOS
	H06z0	Chest infection NOS
	H06z1	Lower resp tract infection

Clinical code group	Read code	Description
Sore Throat	1C8..	Nasal symptoms
	1C9..	Sore throat symptom
	1C92.	Has a sore throat
	1C9Z.	Sore throat symptom NOS
	1CB3.	Throat pain
	2DC3.	Inflamed throat
	H00..	Acute nasopharyngitis
	H02..	Acute pharyngitis
	H024.	Acute viral pharyngitis
	H02z.	Acute pharyngitis NOS
	H050.	Acute laryngopharyngitis
	H121.	Chronic pharyngitis
	H1212	Granular pharyngitis
	R041.	[D]Throat pain
Taste and Smell	1924.	Loss of taste
	192A.	Bad taste in mouth
	1B45.	C/O - loss of smell sense
	2BP3.	O/E - anosmia
	R0110	Anosmia
	R0112	[D]Parageusia
	R011z	[D]Smell or taste disorder NOS
	ZV415	[V]Problem with smell or taste
Tingling	1B41.	Has pins and needles
	1B43.	Has tingling sensation
	1B44.	Has numbness
	1B442	Numbness of limbs
	1B46.	C/O paraesthesia
	1B47.	Transient paraesthesia
	1B48.	Burning feet
	29B4.	O/E - hyperaesthesia present
	2G2D.	Numbness of hand
	C2623	Burning feet syndrome
	F351.	Meralgia paraesthetica
	R0201	[D]Burning of skin
	R0203	[D]Tingling of skin
	R0204	[D]Hyperaesthesia
	R0206	[D]Numbness
	R0207	[D]Paraesthesia
	R0209	[D]Allodynia
Tinnitus	1C2..	Tinnitus symptoms
	F583.	Tinnitus
	F583z	Tinnitus NOS
Viral Infection Influenza	16L..	Influenza-like symptoms
	A79z.	Viral infection NOS
	H051.	Acute upper respiratory tract infection

Table S3: British National Foundry (BNF) sub-paragraph codes

BNF Chapter	BNF Section	BNF Sub-paragraph	BNF Sub-paragraph code
Cardiovascular System	Beta-adrenoceptor blocking drugs	Beta-adrenoceptor blocking drugs	204000
	Hypertension and heart failure	Alpha-adrenoceptor blocking drugs	205040
		Angiotensin-converting enzyme inhibitors	205051
	Nitrates, calcium-channel blockers & other antianginal drugs	Nitrates	206010
		Calcium-channel blockers	206020
	Anticoagulants and protamine	Parenteral anticoagulants	208010
		Oral anticoagulants	208020
	Antiplatelet drugs	Antiplatelet drugs	209000
	Lipid-regulating drugs	Lipid-regulating drugs	212000
Respiratory System	Bronchodilators	Selective beta(2)-agonists	301011
		Compound bronchodilator preparations	301040
	Corticosteroids (respiratory)	Corticosteroids (respiratory)	302000
	Cromoglycate, leukotriene and phosphodesterase type-4 inhibitors	Leukotriene receptor antagonists	303020
	Antihistamines, hyposensitisation and allergic emergencies	Antihistamines	304010
	Cough preparations	Cough suppressants	309010
		Expectorant and demulcent cough preparations	309020
	Systemic nasal decongestants	Systemic nasal decongestants	310000
Infections	Antibacterial drugs	Benzylpenicillin and phenoxymethylpenicillin	501011
		Tetracyclines	501030
		Macrolides	501050
	Antiviral drugs	Herpes simplex and varicella-zoster	503021
		Coronavirus	503060
Endocrine System	Corticosteroids (endocrine)	Replacement therapy	603010
Nutrition and Blood	Anaemias and some other blood disorders	Oral iron	901011
	Vitamins	Nicotinamide (B7)	906022
		Vitamin D	906040
		Multivitamin preparations	906070

Supplementary methods

Detailed description of development of the operational definition for long COVID

To develop an operational definition that could be used to identify individuals as having long COVID or not, we used matched analysis to identify individual indicators of long COVID and then investigated how those indicators cluster to form one or more phenotypes for long COVID, as depicted in Figure S1**: Schematic of the methods used to create an operational definition of long COVID**.

Preparing the matched cohort

We began by preparing a matched cohort consisting of pairs of individuals with positive and negative RT-PCR test results for SARS-CoV-2, and with the same propensity to receive a positive RT-PCR test in a given month.

We used time-varying matching in month-long intervals from 1 March 2020 until 30 April 2022 (when widespread RT-PCR testing ended in Scotland). In each time period, we matched individuals whose first positive RT-PCR test was recorded during the period (exposed group) to individuals whose first negative RT-PCR test was recorded during the period (including only those individuals who had not previously tested positive) (control group). Individuals were matched on their propensity to test positive for SARS-CoV-2 during the period under investigation. We used nearest neighbour matching on propensity scores with a calliper of 0.8 standard deviations, coupled with exact matching on week of test and age in years (we used individual years of age up to 79, then two-year age bands from 80-89, five-year bands for 90-99, and a single band for those aged 100 or older). Individuals with a positive RT-PCR test were eligible to be used as controls up until the date at which they tested positive. In the event that an exposed case had more than one candidate control, the control case whose propensity score was closest to the exposed case's score was selected. We included the restriction that each individual could be used as a control case no more than once.

Our propensity score model included the following predictors: splines in age (with three degrees of freedom); sex; SIMD quintile; six-fold urban-rural classification; local authority of residence; household size (of which 12.1% was mean imputed); number of COVID-19 vaccine doses received up to 14 days before the test date used for matching; and number of RT-PCR tests taken by the RT-PCR test date used for matching; presence or absence of each of the clinical risk groups listed in Table 2 (reflecting the subset of predictors used in the Q-COVID algorithm[1] for which Scottish data is available); splines in BMI (with three degrees of freedom), of which 61.8% was imputed using a bidirectional stepwise regression model; binary indicators of individuals' status as immunosuppressed, recommended to shield, or having been hospitalised or admitted to an ICU in the 12 months before testing.

As a sensitivity test, we repeated the propensity score matching using a second control group consisting of individuals who had neither a positive nor a negative RT-PCR test by the beginning of the month under investigation. Control group members' lack of a test date necessitated that we omit exact matching on week of test in this version of the analysis. It also precluded the inclusion of number of RT-PCR tests taken as a predictor in our propensity score model (because all controls had zero tests and all members of the exposed group had one or more tests, by definition). Controls were assigned a pseudo test date (selected at random from the range of dates within the month under investigation) to allow for calculation of the subset of predictors that were dependent on individuals' test dates (including number of COVID-19 vaccine doses received up to 14 days before testing and hospitalisations and ICU admissions during the 12 months prior to testing).

54.4% of positive cases were matched to a control with a negative RT-PCR test, and 84.4% of positive cases were matched to a control who had not yet tested. Covariate balance plots (Figure S2: Covariate plots used to assess balance in the matched samples) were used to confirm the adequacy of the matching.

Matched analysis

Matched pairs were jointly censored if: either individual died before the end of the follow up period; if the control tested positive for COVID-19; or if the exposed individual was reinfected, as identified by a second positive RT-

PCR test at least 42 days after the initial positive test (a cut-off of 42 days was selected to allow for persistence of viral material from the original infection, following advice provided by Public Health Scotland).

We fitted individual Poisson regression models to estimate adjusted rate ratios (aRR) for exposed cases, relative to control cases for each of our dependent variables. The dependent variables captured counts of various clinical interactions recorded in EHR within 4-12 weeks and >12-26 weeks of the exposed case's test date in each matched pair. The clinical interactions we considered included: 45 grouped clinical codes (reflecting symptoms and diagnoses recorded in primary care records, listed in **Error! Reference source not found.**); 27 newly dispensed categories of prescriptions (whereby 'newly dispensed' refers to prescriptions that had not been dispensed in the 12 months prior to the test date used for matching, listed in Table S3: British National Foundry (BNF) sub-paragraph codes), and seven indicators of health service use (including counts of: GP visits, hospital admissions, outpatient attendances for respiratory conditions, A&E visits, out of hours encounters, intensive care unit (ICU) admissions, and NHS 24 telehealth interactions).

Each model included an offset to account for variation in number of days of follow-up. To adjust for any residual imbalance remaining after matching, all predictors used in the propensity score estimation were included as covariates. The Quasi-Poisson variant of Poisson regression was used to adjust for the possibility of overdispersion. We adjusted p-values to reduce the false discovery rate, using Benjamini and Hochberg's method.[2] Dependent variables that were recorded in fewer than five exposed or control cases' EHR were removed from the analysis.

Figures Figure **S**3: Rate ratios of clinical interactions for individuals with a positive RT-PCR test, relative to individuals with a negative RT-PCR test 4-12 weeks following testing**-**Figure **S**4: Rate ratios of clinical interactions for individuals with a positive RT-PCR test, relative to individuals with a negative RT-PCR test >12-26 weeks following testing present the results of the matched analysis, comparing individuals with a positive RT-PCR test to those with a negative RT-PCR test, 4-12 weeks and >12-26 weeks following positive cases' test date, respectively. Figures Figure **S**5**: Rate ratios of clinical interactions for individuals with a positive RT-PCR test, relative to individuals who have not yet taken an RT-PCR test, 4-12 weeks following positive cases' test dates-**Figure **S**6**: Rate ratios of clinical interactions for individuals with a positive RT-PCR test, relative to individuals who have not yet taken an RT-PCR test, >12-26 weeks following positive cases' test dates** present the equivalent results comparing positive cases to controls who had not yet taken an RT-PCR test. All clinical interactions that occurred at a significantly higher rate in the analysis comparing the exposed group to controls with a negative RT-PCR test also occurred at a significantly higher rate relative to controls who had not yet taken an RT-PCR test (adjusted-$p < .05$).

As a sensitivity test, we repeated the matched analysis, stratified by periods when the wild, Alpha, Delta and Omicron SARS-CoV-2 variants were dominant. The results presented in Figures Figure S7**: Rate ratios of symptoms, diagnoses and health service use for individuals with a positive RT-PCR test, relative to those** with a negative RT-PCR, 4-12 **weeks following positive cases' test dates, stratified by variant period**-Figure S10: **Rate ratios of dispensed prescriptions for individuals with a positive RT-PCR test, relative to those** with a negative RT-PCR, >12-26 **weeks following positive cases' test dates, stratified by variant period** reveal similar patterns across the earlier periods and a reduction in significantly higher rates of clinical interactions in the 4-12 and >12-26 weeks following testing for positive cases during the Omicron period. Significantly lower rates of the 'Covid' indicator during the Delta and Omicron periods in the 4-12 and >12-26 weeks following testing may indicate that fewer positive cases were captured in RT-PCR testing data during these periods.

Cluster analysis

To identify one or more phenotypes for long COVID, we performed cluster analysis on the indicators that occurred at a significantly higher rate among individuals with a positive RT-PCR test relative to those with a negative RT-PCR test in the 4-12 or >12-26 weeks after testing. We tested for clusters of indicators among exposed cases only. Counts of each indicator were normalised prior to clustering. We performed hierarchical clustering using Gower's

measure of distance to calculate the dissimilarity between indicators, coupled with Ward's linkage method. We calculated Dunn's Index and the average silhouette widths for two to twelve clusters to identify the optimal number of clusters in terms of internal validity. We then plotted dendrograms colour-coded by cluster, shown in Figures **Figure S**11**:** Clusters of long COVID indicators at 4-12 weeks identified using Hierarchical Clustering-Figure **S**12: Clusters of long COVID indicators at >12-26 weeks identified using Hierarchical Clustering. For sensitivity, we repeated the analysis using k-medoids clustering (also known as partition around medoids (PAM)). As before, we calculated Dunn's Index and the average silhouette widths for two to twelve clusters to identify the optimal number of clusters in terms of internal validity. The optimal cluster solutions are presented in Tables **Error! Not a valid bookmark self-reference.**-Table **S**5**: Clusters of long COVID indicators at >12-26 weeks identified using Partition Around Medoids**.

As discussed in the main text, imbalance in the frequencies of clinical outcomes that are and are not automatically coded in EHR, coupled with sparse recording of outcomes that are not automatically coded, necessitated that we adopt an alternative approach to developing our operational definition, informed by clinical practice (Figure 2).

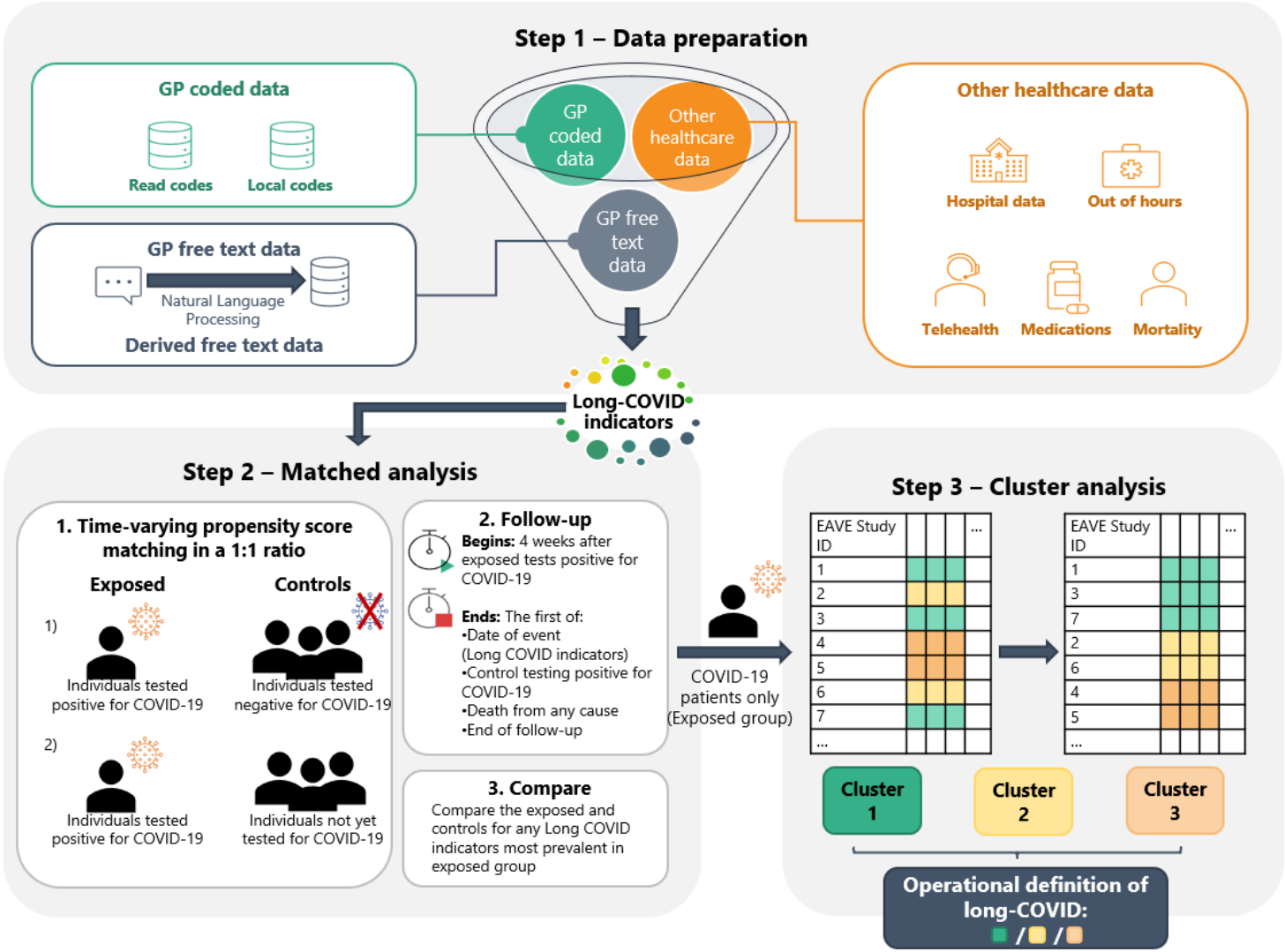


Figure S1: Schematic of the methods used to create an operational definition of long COVID

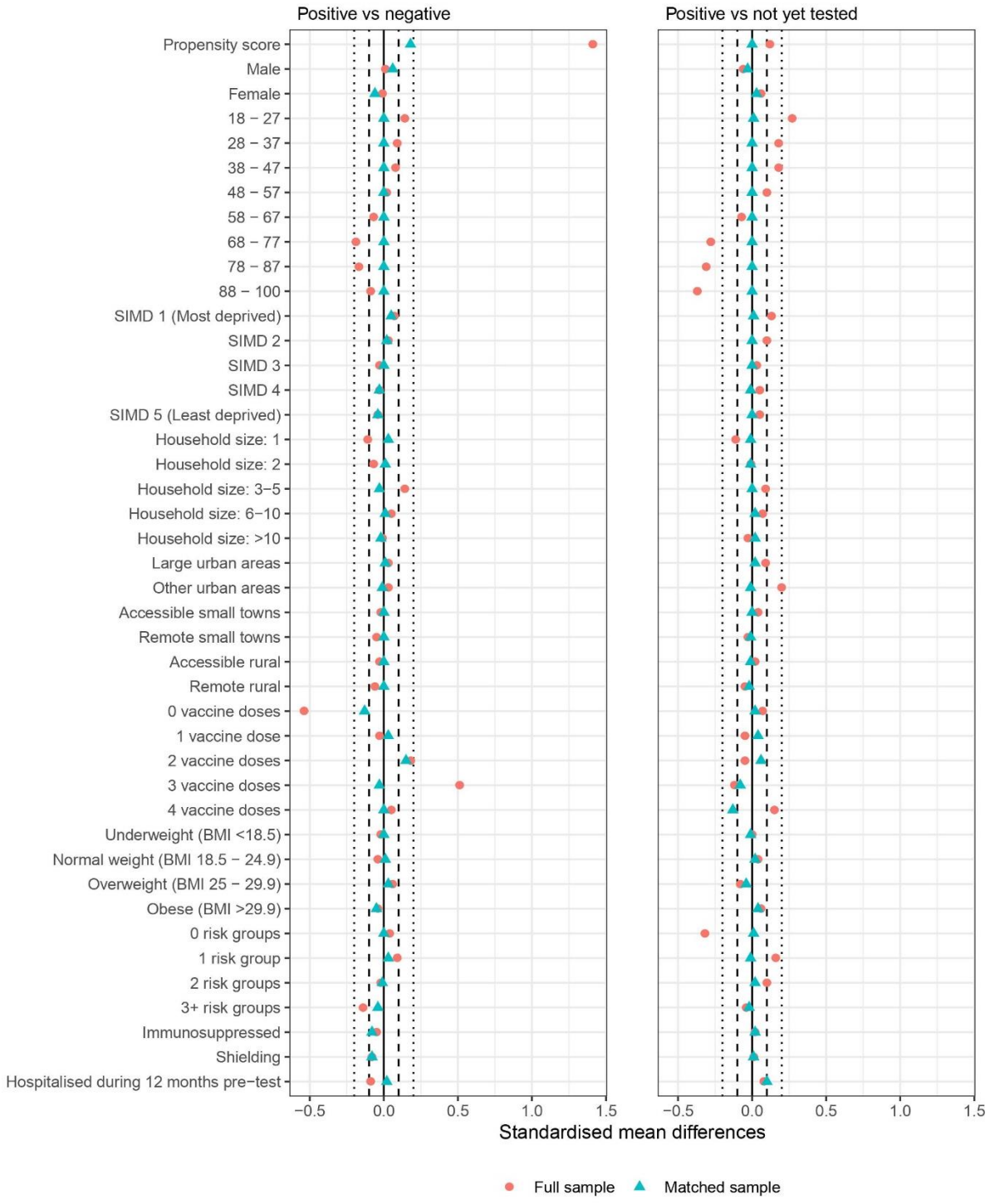
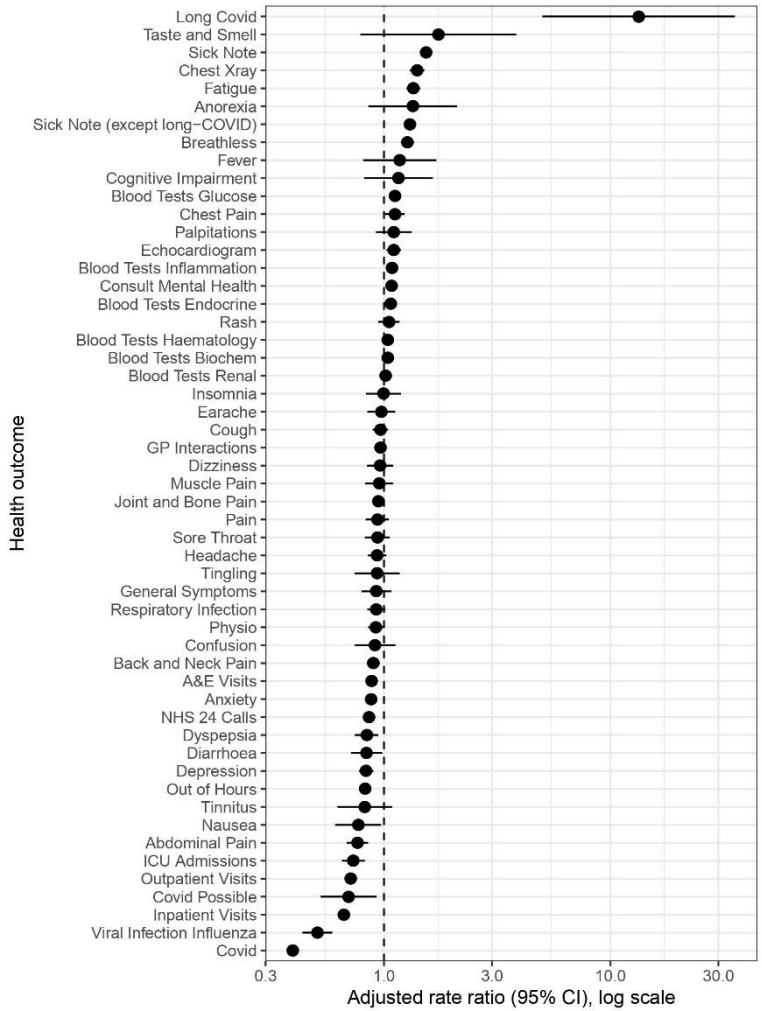
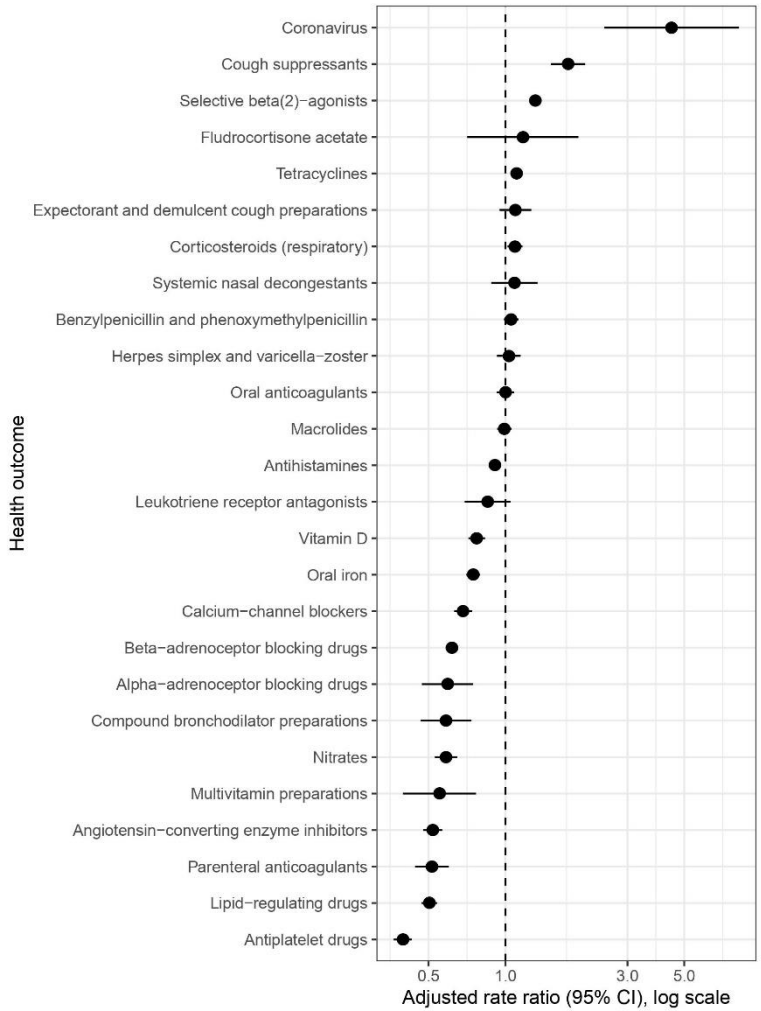


Figure S2: Covariate plots used to assess balance in the matched samples
Standardised mean differences between exposed and control groups in the full cohort and the matched sample are shown for each control group (individuals with a negative RT-PCR test and individuals who have not yet tested). Positive standardised mean differences indicate larger means in the exposed group relative to the control group. Points between the dashed lines indicate that values for the exposed and control groups are within 0.1 standard mean differences. Points between the dotted lines indicate that values for the exposed and control groups are within 0.2 standard mean differences.



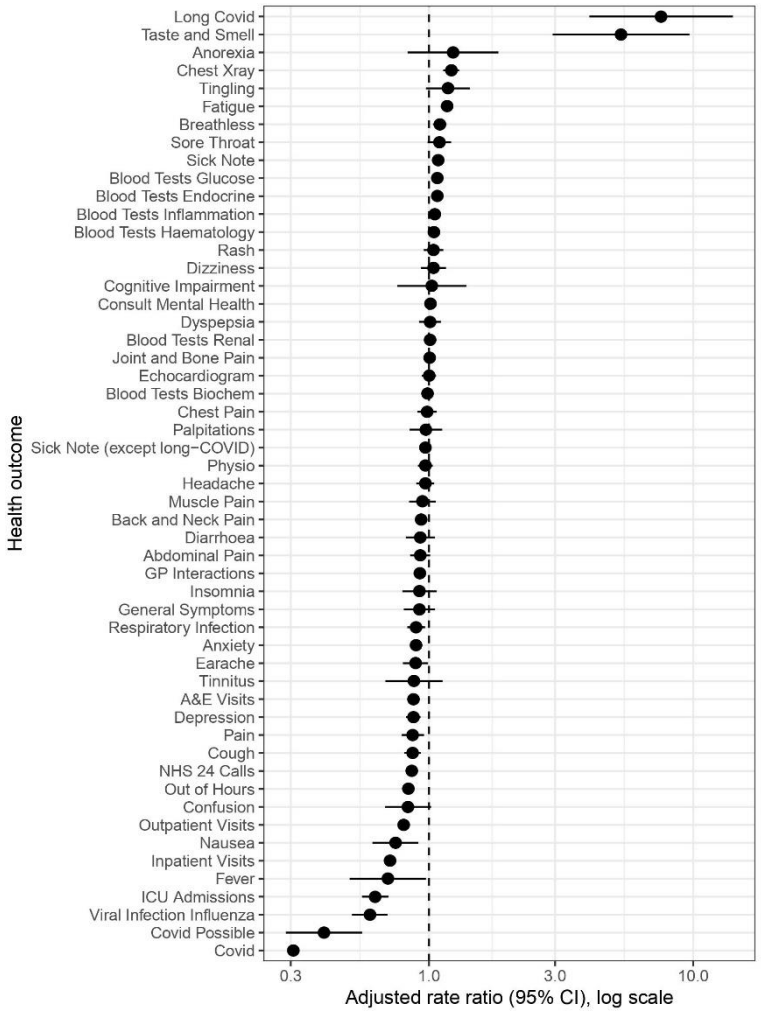
A.



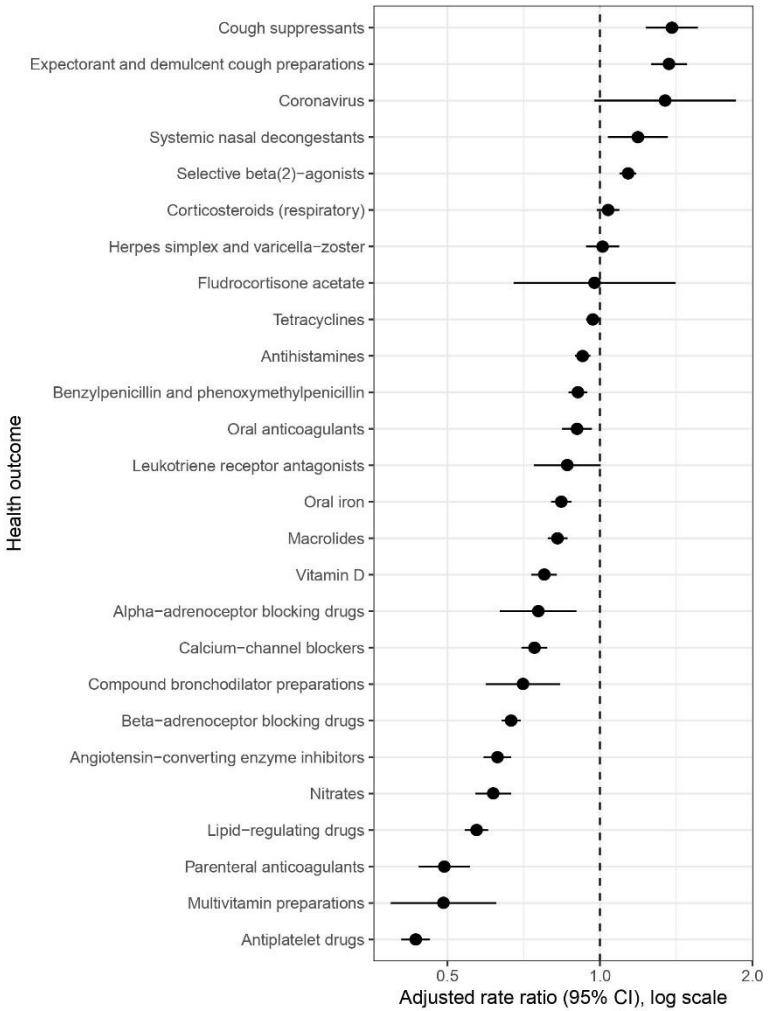
B.

Figure S3: Rate ratios of clinical interactions for individuals with a positive RT-PCR test, relative to individuals with a negative RT-PCR test 4-12 weeks following testing

Panel A presents rate ratios for the 45 grouped clinical codes (reflecting symptoms and diagnoses recorded in primary care records) and the seven indicators of health service use. Panel B presents rate ratios for the 27 newly dispensed categories of prescriptions (whereby 'newly dispensed' refers to medications that had not been dispensed in the 12 months prior to testing). Each point represents an estimate from a separate Poisson regression model. Regressions were run on the matched sample. Controls for all variables listed in Table 1 were included in each regression model. 95% confidence intervals are shown.



A.



B.

Figure S4: Rate ratios of clinical interactions for individuals with a positive RT-PCR test, relative to individuals with a negative RT-PCR test >12-26 weeks following testing

Panel A presents rate ratios for the 45 grouped clinical codes (reflecting symptoms and diagnoses recorded in primary care records) and the seven indicators of health service use. Panel B presents rate ratios for the 27 newly dispensed categories of prescriptions (whereby 'newly dispensed' refers to medications that had not been dispensed in the 12 months prior to testing). Each point represents an estimate from a separate Poisson regression model. Regressions were run on the matched sample. Controls for all variables listed in Table 1 were included in each regression model. 95% confidence intervals are shown.

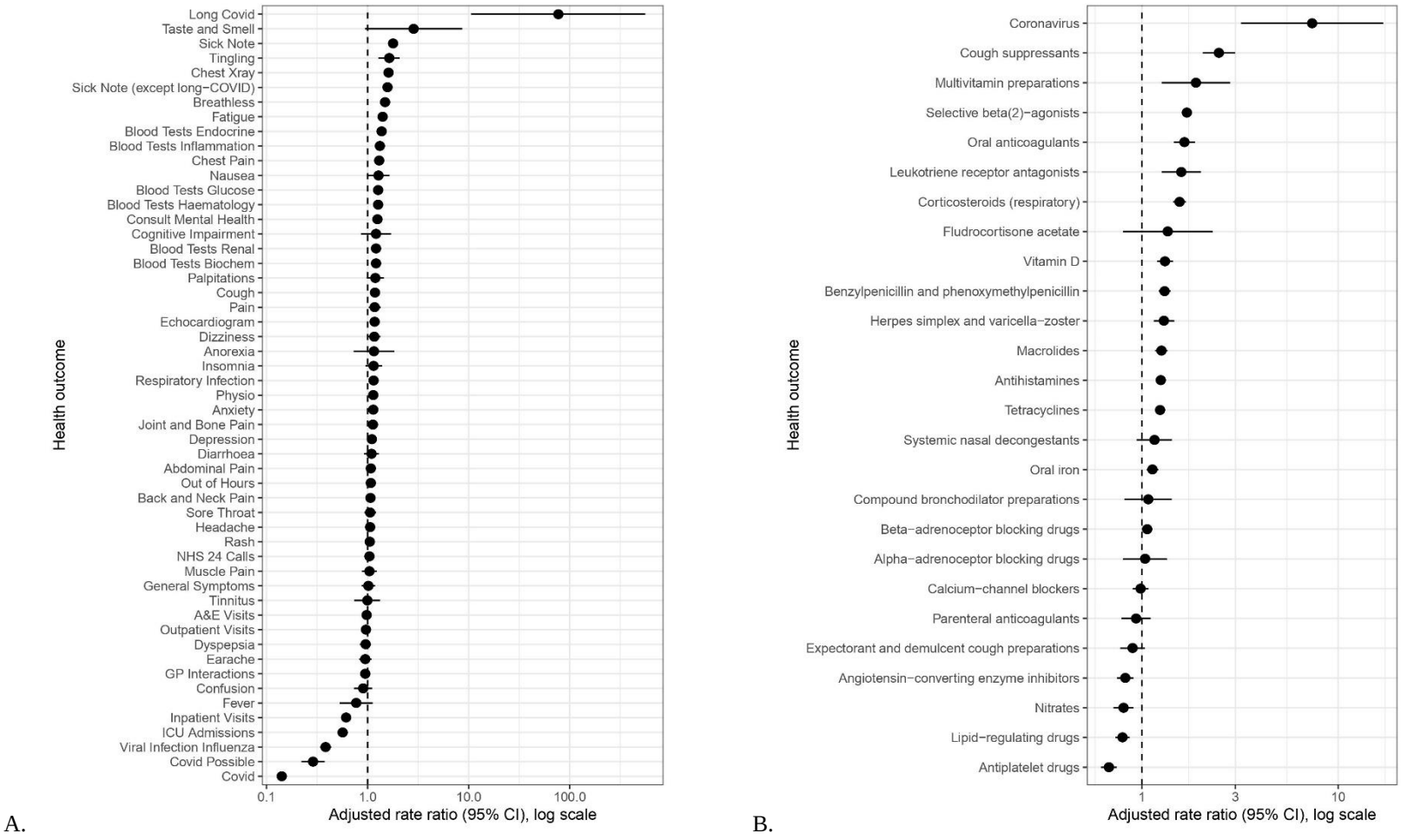


Figure S5: Rate ratios of clinical interactions for individuals with a positive RT-PCR test, relative to individuals who have not yet taken an RT-PCR test, 4-12 weeks following positive cases' test dates

Panel A presents rate ratios for the 45 grouped clinical codes (reflecting symptoms and diagnoses recorded in primary care records) and the seven indicators of health service use. Panel B presents rate ratios for the 27 newly dispensed categories of prescriptions (whereby 'newly dispensed' refers to medications that had not been dispensed in the 12 months prior to testing). Each point represents an estimate from a separate Poisson regression model. Regressions were run on the matched sample. Controls for all variables listed in Table 1 were included in each regression model. 95% confidence intervals are shown.

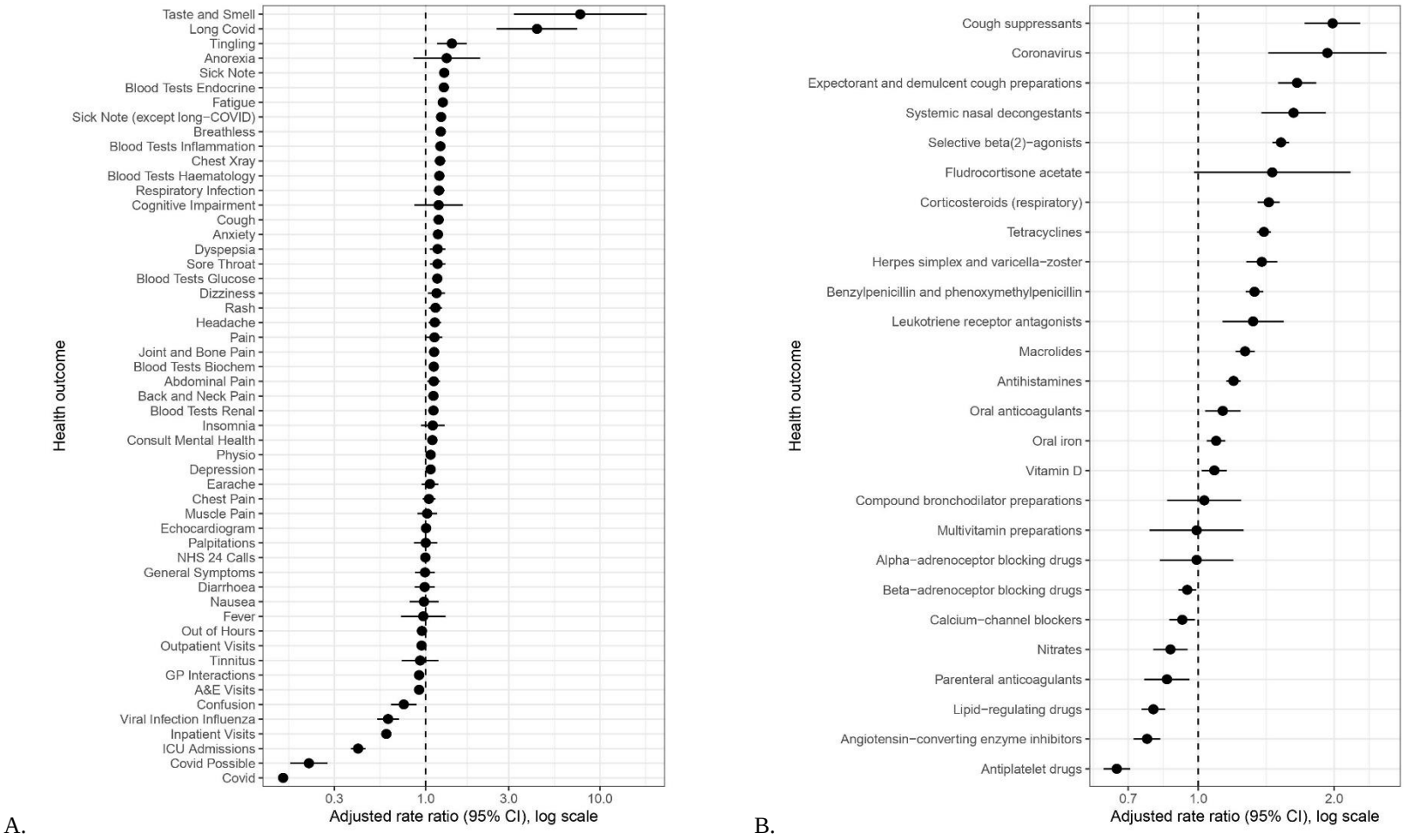


Figure S6: Rate ratios of clinical interactions for individuals with a positive RT-PCR test, relative to individuals who have not yet taken an RT-PCR test, >12-26 weeks following positive cases' test dates

Panel A presents rate ratios for the 45 grouped clinical codes (reflecting symptoms and diagnoses recorded in primary care records) and the seven indicators of health service use. Panel B presents rate ratios for the 27 newly dispensed categories of prescriptions (whereby 'newly dispensed' refers to medications that had not been dispensed in the 12 months prior to testing). Each point represents an estimate from a separate Poisson regression model. Regressions were run on the matched sample. Controls for all variables listed in Table 1 were included in each regression model. 95% confidence intervals are shown.

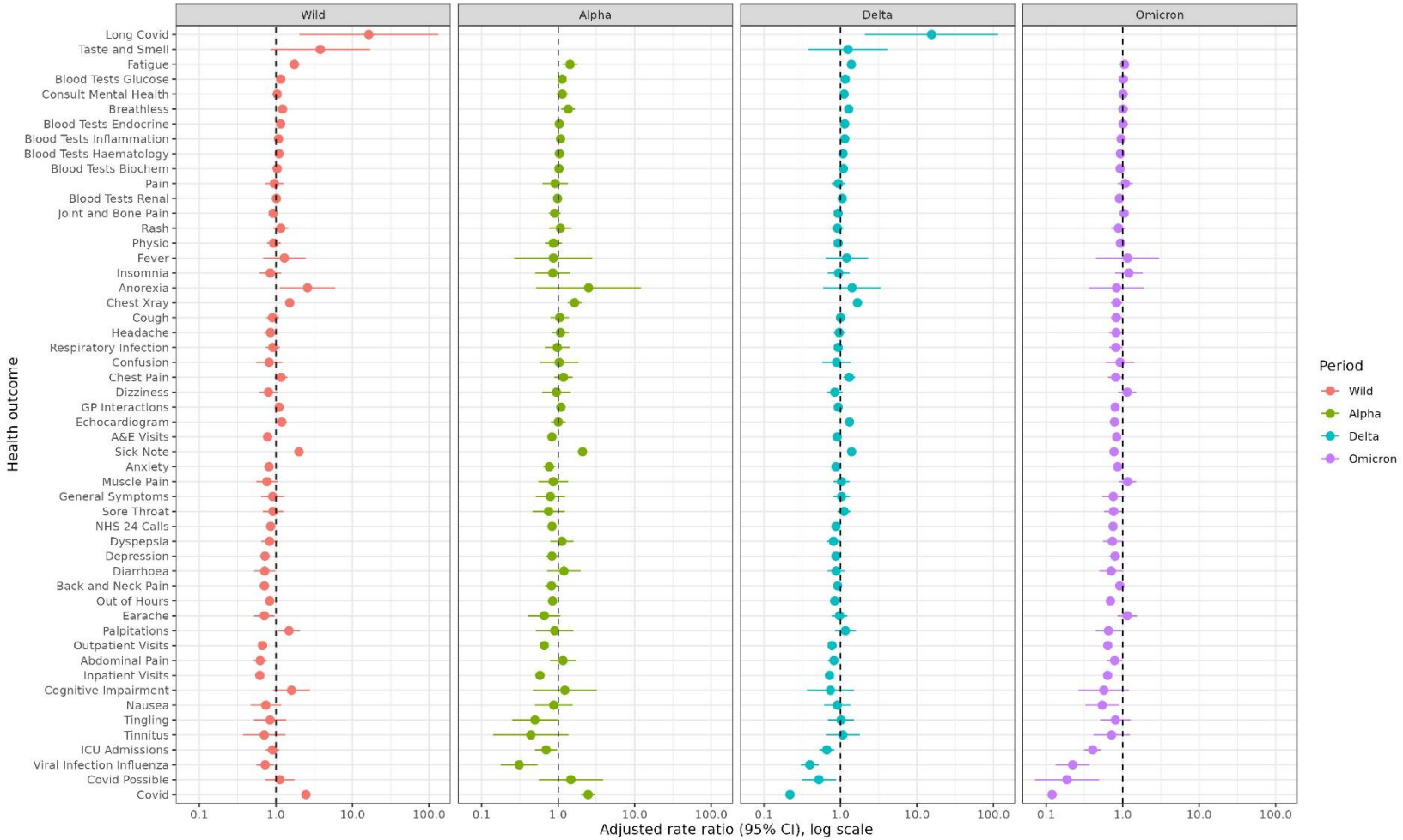


Figure S7: Rate ratios of symptoms, diagnoses and health service use for individuals with a positive RT-PCR test, relative to those with a negative RT-PCR, 4-12 weeks following positive cases' test dates, stratified by variant period

The plot shows rate ratios for the 45 grouped clinical codes (reflecting symptoms and diagnoses recorded in primary care records) and the seven indicators of health service use. Analysis is stratified according to the variant period in which tests were taken. Variant periods reflect the periods when the following strains of SARS-CoV-2 represented more than 60% of sequenced cases: wild (1 March 2020 to 10 January 2021), Alpha (11 January 2021 to 9 May 2021), Delta (24 May 2021 to 28 November 2021), Omicron (20 December 2021 to 30 April 2021). Each point represents an estimate from a separate Poisson regression model. Missing point estimates occur where there were too few observations for the model to converge. Regressions were run on the matched sample. Controls for all variables listed in Table 1 were included in each regression model. 95% confidence intervals are shown.

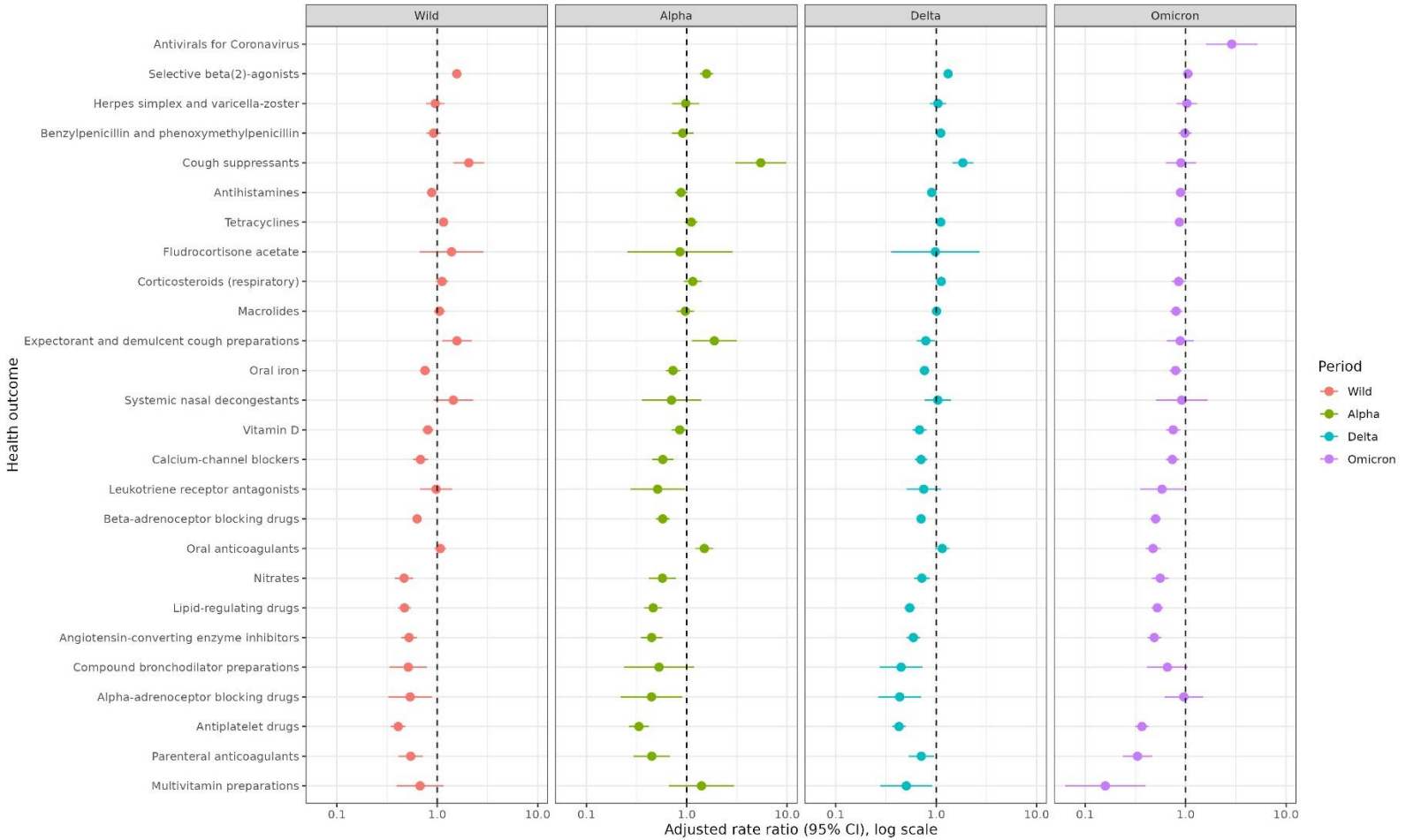


Figure S8: Rate ratios of dispensed prescriptions for individuals with a positive RT-PCR test, relative to those with a negative RT-PCR, 4-12 weeks following positive cases' test dates, stratified by variant period

The plot shows rate ratios for the 27 newly dispensed categories of prescriptions (whereby 'newly dispensed' refers to medications that had not been dispensed in the 12 months prior to testing). Analysis is stratified according to the variant period in which tests were taken. Variant periods reflect the periods when the following strains of SARS-CoV-2 represented more than 60% of sequenced cases: wild (1 March 2020 to 10 January 2021), Alpha (11 January 2021 to 9 May 2021), Delta (24 May 2021 to 28 November 2021), Omicron (20 December 2021 to 30 April 2021). Each point represents an estimate from a separate Poisson regression model. Regressions were run on the matched sample. Controls for all variables listed in Table 1 were included in each regression model. 95% confidence intervals are shown.

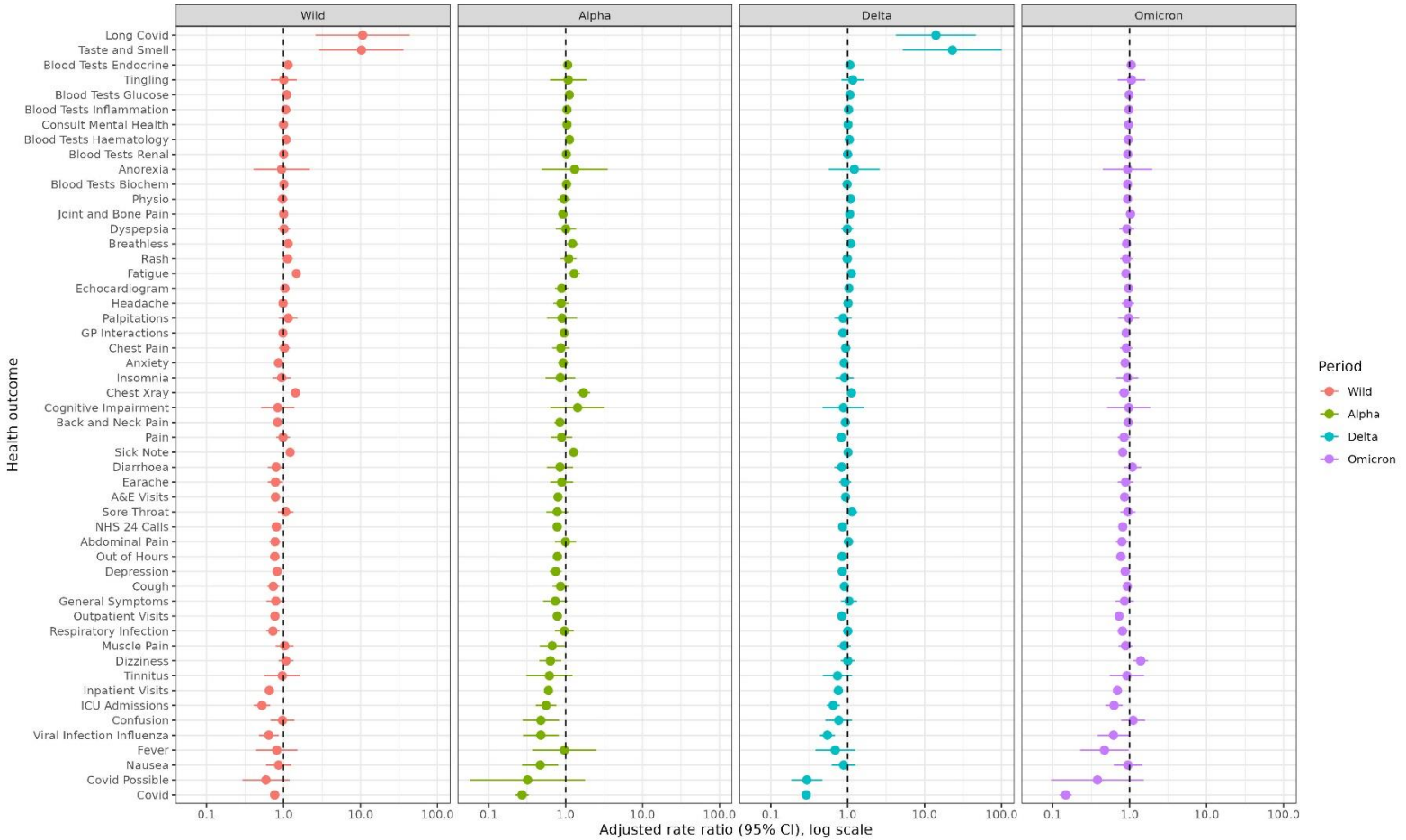


Figure 9: Rate ratios of symptoms, diagnoses and health service use for individuals with a positive RT-PCR test, relative to those with a negative RT-PCR, >12-26 weeks following positive cases' test dates, stratified by variant period

The plot shows rate ratios for the 45 grouped clinical codes (reflecting symptoms and diagnoses recorded in primary care records) and the seven indicators of health service use. Analysis is stratified according to the variant period in which tests were taken. Variant periods reflect the periods when the following strains of SARS-CoV-2 represented more than 60% of sequenced cases: wild (1 March 2020 to 10 January 2021), Alpha (11 January 2021 to 9 May 2021), Delta (24 May 2021 to 28 November 2021), Omicron (20 December 2021 to 30 April 2021). Each point represents an estimate from a separate Poisson regression model. Missing point estimates occur where there were too few observations for the model to converge. Regressions were run on the matched sample. Controls for all variables listed in Table 1 were included in each regression model. 95% confidence intervals are shown.

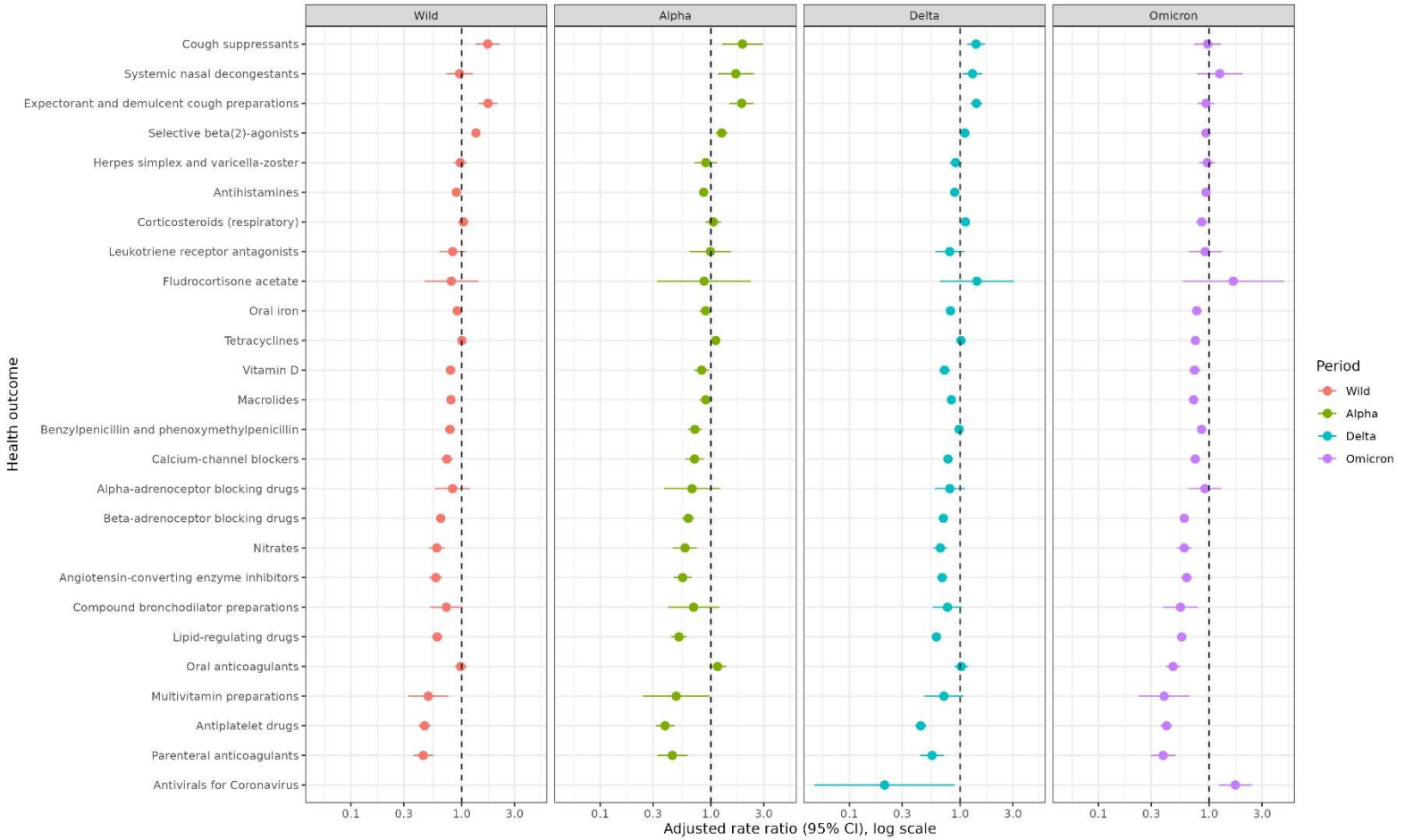


Figure S10: Rate ratios of dispensed prescriptions for individuals with a positive RT-PCR test, relative to those with a negative RT-PCR, >12-26 weeks following positive cases' test dates, stratified by variant period

The plot shows rate ratios for the 27 newly dispensed categories of prescriptions (whereby 'newly dispensed' refers to medications that had not been dispensed in the 12 months prior to testing). Analysis is stratified according to the variant period in which tests were taken. Variant periods reflect the periods when the following strains of SARS-CoV-2 represented more than 60% of sequenced cases: wild (1 March 2020 to 10 January 2021), Alpha (11 January 2021 to 9 May 2021), Delta (24 May 2021 to 28 November 2021), Omicron (20 December 2021 to 30 April 2021). Each point represents an estimate from a separate Poisson regression model. Missing point estimates occur where there were too few observations for the model to converge. Regressions were run on the matched sample. Controls for all variables listed in Table 1 were included in each regression model. 95% confidence intervals are shown.

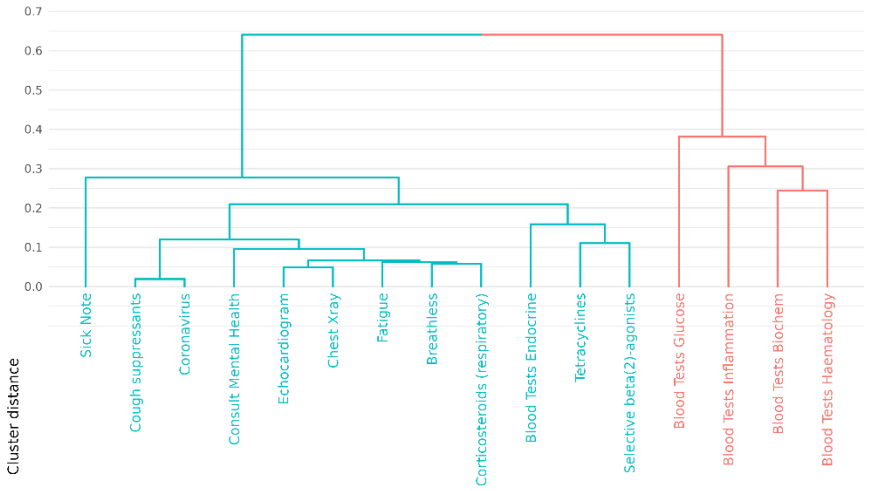


Figure S11: Clusters of long COVID indicators at 4-12 weeks identified using Hierarchical Clustering
Hierarchical Clustering was performed on indicators that occur at a significantly higher rate among individuals with a positive SARS-CoV-2 test, relative to individuals with a negative SARS-CoV-2 during the 4-12 weeks after testing. Gower's distance method and Ward's agglomeration method were used. Two clusters were selected based on the optimal silhouette width for 2-12 clusters.

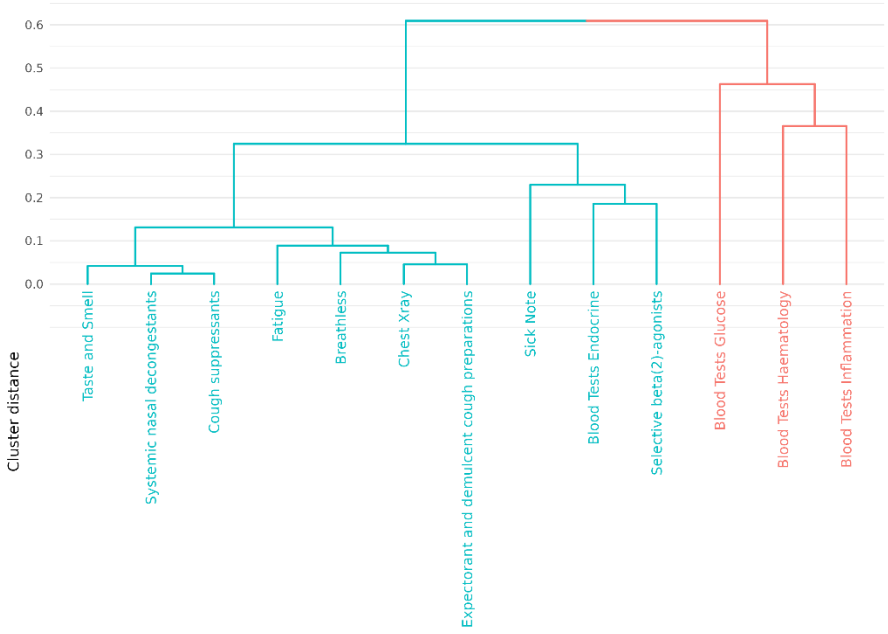


Figure S12: Clusters of long COVID indicators at >12-26 weeks identified using Hierarchical Clustering
Hierarchical Clustering was performed on indicators that occur at a significantly higher rate among individuals with a positive SARS-CoV-2 test, relative to individuals with a negative SARS-CoV-2 during the >12-26 weeks after testing. Gower's distance method and Ward's agglomeration method were used. Two clusters were selected based on the optimal silhouette width for 2-12 clusters

Table S4: Clusters of long COVID indicators at 4-12 weeks identified using Partition Around Medoids

Cluster	Outcome
1	Sick Note
2	Chest Xray
	Breathless
	Fatigue
	Echocardiogram
	Blood Tests Endocrine
	Consult Mental Health
	Corticosteroids (respiratory)
	Tetracyclines
	Selective beta(2)-agonists
	Cough suppressants
	Coronavirus
3	Blood Tests Glucose
4	Blood Tests Inflammation
5	Blood Tests Haematology
6	Blood Tests Biochem

Indicators that occurred at a significantly higher rate among individuals with a positive SARS-CoV-2 test, relative to individuals with a negative SARS-CoV-2 test in the 4-12 weeks after testing were clustered using Partition Around Medoids (k-medoids) clustering using Gower's distance measure. Six clusters were selected based on the optimal value of Dunn's index.

Table S5: Clusters of long COVID indicators at >12-26 weeks identified using Partition Around Medoids

Cluster	Outcome
1	Blood Tests Haematology
2	Blood Tests Inflammation
3	Blood Tests Endocrine
	Sick Note
	Breathless
	Fatigue
	Chest Xray
	Taste and Smell
	Selective beta(2)-agonists
	Systemic nasal decongestants
	Expectorant and demulcent cough preparations
	Cough suppressants
4	Blood Tests Glucose

Indicators that occur at a significantly higher rate among individuals with a positive SARS-CoV-2 test, relative to individuals with a negative SARS-CoV-2 test in the >12-26 weeks after testing were clustered using Partition Around Medoids (k-medoids) clustering with Gower's distance measure. Four clusters were selected to optimise Dunn's index.

Patient and public involvement with this study

Table S66: GRIPP2 reporting checklist (short form)

Section and topic	Item	Reported on page No
1: Aim	Report the aim of PPI in the study	S33
2: Methods	Provide a clear description of the methods used for PPI in the study	S33
3: Study results	Outcomes—Report the results of PPI in the study, including both positive and negative outcomes	S33-34
4: Discussion and conclusions	Outcomes—Comment on the extent to which PPI influenced the study overall. Describe positive and negative effects	S34-35
5: Reflections/critical perspective	Comment critically on the study, reflecting on the things that went well and those that did not, so others can learn from this experience	S35

PPI=patient and public involvement

Aim

The aims of patient and public involvement (PPI) in this study were to: (1) embed patient and public perspectives and information needs into project decision-making; (2) ensure that the experiences of people with Long Covid were incorporated into the study design; and (3) contribute to shared best practice in PPI.

Methods

This study uses routinely collected health data from the "Early Pandemic Evaluation and Enhanced Surveillance of COVID-19" (EAVE II) platform, including clinical codes, free text from patient notes, sick notes and prescribing. The initial research bid was reviewed in October 2020 by our Lay Co-Investigator (Weatherill), selected from a wide pool of regular patient and public contributors due to his PPI experience and understanding of healthcare data in Scotland via co-leadership of the EAVE II Public Advisory Group (PAG, n=15). The PPI Coordinator (Woolford) recruited two patient partners (Batchelor, White) in July 2021 from the Long Covid Scotland Action Group, on the basis of research interest and lived experience of Long Covid.

The resulting PPI Team have been involved in co-producing a PPI Strategy; steering the project; commenting on the analysis protocol; designing and releasing a patient survey to inform analysis interpretation; reviewing plain English summaries of project outputs; assisting in the development of public- and policy-facing documents released in conjunction with the outputs of this study, and authoring this GRIPP2 Appendix.

This work was carried out remotely, either using videoconferencing (Zoom, with minutes produced from each recording) or asynchronously via email. Public members of the PPI Team were paid for time and expertise shared in line with National Institute for Health and Care Research (NIHR) guidelines,[3] with appropriate paperwork issued to prevent compromise of any state financial support received. Role Descriptions, Terms of Reference and PPI Objectives and Timelines were co-produced and agreed by the PPI Team shortly after recruitment.

Results

The newly formed PPI team decided on aims for the project as described above, and deliverables throughout the research cycle to achieve these, summarised in **Table S7**7S7.

Table S77: Results of PPI

Area of research cycle	Summary of deliverables
Grant development	Appoint Lay Co-Investigator and comment on grant application.
Undertaking project	Collaborate with Long Covid Scotland; co-produce PPI Strategy and Terms of Engagement; provide induction and statistical methods training for project.
Design	Review analysis protocol; support design and release of survey gauging symptoms and impact of Long Covid on patients in Scotland; continue to question and comment on design development at Steering Group and PPI meetings.
Analysis and interpretation	Share results from Long Covid Scotland with analysts to inform interpretation; design and carry out consultation with people with Long Covid to select features for prediction model from patient perspective.
Dissemination	Review public-facing outputs to produce plain English resources an identify potential questions; collaborate with staff to provide written contributions for academic publications.
Implementation	Provide steer on appropriate messaging and content to be released in policy briefing(s) and in any supporting media materials.
Evaluation	Evaluate PPI element of project in final stages; share this work by means of a PPI report or paper.

Following this, the PPI team met with the primary analyst at the project's design phase to discuss the analysis protocol. Potential shortcomings, including a lack of data from private referrals for Long Covid, were highlighted as part of this involvement.

To further the collaborative design of the project, a survey of Long Covid Scotland members led by White was reviewed by experienced clinical and analytical members of the project team before release, particularly in relation to questions on demographic factors and symptoms. The survey was designed to give a richer understanding of the physical, mental, relational and financial impacts of Long Covid on members of the group, beyond what could be established through routinely collected data. The findings, based on 222 in-depth responses from across Scotland, were shared by CW with our team and via the Scottish Parliament's Cross-Party Group (CPG) for Long Covid.[4] They are available through the 2022 report 'Hearing Our Voices - Long Covid: The Impact on Our Lives',[5] and have been used to inform refinement of the analysis design and results interpretation. This includes confirmation that survey symptoms were consistent with operational definition results, but also acknowledgment of the challenge of capturing commonly reported symptoms such as cognitive dysfunction in clinical codes.

The PPI Team have attended a total of 14 Steering Group meetings and 4 dedicated PPI meetings, to steer and comment on project development. Key conversations included discussions of inclusion criteria; how operational definitions can help to address the issue of identifying Long Covid patients who may not have had access to testing; whether different SARS-CoV-2 variants pose a higher Long Covid risk; patient interactions with GPs and symptom curation; lobbying for more timely data access; and discussing the timing and sensitivity of disseminating project results to policymakers and the public.

As new data and preliminary results emerged, it became clear that our initial deliverable of involving a wider group of Long Covid Scotland members in interpreting the results of cluster analysis data would be impractical, as there were insufficient numbers of patients with multiple codes in the cohort to generate clear clusters. Instead, we have committed to involving patient contributors in feature selection for the prediction model, alongside clinical and analytical experts.

As part of ongoing conversations with our public contributors on dissemination of results, we have agreed to collaborate with key policy stakeholders including Public Health Scotland, Scottish Government and the Chief Scientist Office (CSO).

Discussion and conclusions

As part of the development of an operational definition of Long Covid, patient and public contributions have shaped the analysis design, project steering and results interpretation in a significant way. The sharing of lived experience has been of particular benefit to understanding how results derived from routinely collected data fit into the wider context of patient interaction with the healthcare system.

Involvement is also helping to define how we will disseminate the study and contribute to policy implementation in Scotland, for a new condition which continues to impact the lives of many.

Reflections

The longer timeframe of study has allowed for more nuanced input from PPI through Steering Group and PPI meetings, which have represented an effective mechanism for involvement in this project. Multiple, shorter opportunities for input have also been particularly important when collaborating with people living with chronic conditions like Long Covid, which are unpredictable and likely to cause fatigue.

From a coordination perspective, the nature and length of the project has also necessitated more regular input. The PPI Team have had to adapt considerably to changes in project design and management due to data access delays and implications of preliminary results, particularly for cluster analysis. This is an important learning point which has been pronounced in this project due to the complexity and novelty of analysis, involving a new chronic condition, poor clinical coding, and free text analysis.

From a patient perspective, involvement represents a critical aspect of a project exploring a new health condition. By definition, much of the expertise surrounding Long Covid is currently held in lived experience. The PPI activities carried out have allowed for more in-depth understandings of how Long Covid impacts people's lives. They also provided a platform for longer-term knowledge exchange between patient, clinical and analytical experts. The low incidence of multiple codes for a given patient, which influenced the final operational definition of Long Covid in this study, reflects anecdotal patient experience in which consultation conversations may omit, overlook or minimise descriptions of multiple symptoms.

From an analysts' perspective, recruitment of PPI patient partners who were active members of a wider patient group was particularly valuable. Due to their involvement in the Long Covid Scotland Action Group, the patient partners were able to share not only their own experiences, but also those of the broader Long Covid community in Scotland. The PPI Team's work on the patient survey was particularly beneficial, as it provided a formal mechanism for PPI Team members to synthesise the experiences of the wider Long Covid community. Their input at steering group meetings and in the "Hearing our Voices - Long Covid: The Impact on Our Lives" survey report greatly enhanced the project team's ability to make informed analytical decisions and interpret results from a patient-centred perspective.

Table S8: Individuals identified by long COVID outcome measures, stratified by health boards

	Full sample		Any outcome		Long-COVID clinical code		Long-COVID in free text		Long-COVID on sick note		Operational definition	
	N	%	N	%	N	%	N	%	N	%	N	%
Total (% of population)	5,104,198	100.0	90,712	1.8	1,092	0.0	8,368	0.2	14,471	0.3	73,767	1.4
NHS Ayrshire and Arran	273,254	5.4	8,202	9.0	17	1.6	208	2.5	1,675	11.6	7,024	9.5
NHS Borders	83,942	1.6	1,388	1.5	<5	-	68	0.8	299	2.1	1,130	1.5
NHS Dumfries and Galloway	125,852	2.5	2,615	2.9	<5	-	83	1.0	422	2.9	2,285	3.1
NHS Fife	288,467	5.7	6,240	6.9	16	1.5	150	1.8	1,382	9.6	5,165	7.0
NHS Forth Valley	232,562	4.6	5,181	5.7	<5	-	44	0.5	1,564	10.8	4,024	5.5
NHS Grampian	467,235	9.2	6,832	7.5	120	11.0	1,106	13.2	227	1.6	5,730	7.8
NHS Greater Glasgow and Clyde	1,022,047	20.0	24,833	27.4	34	3.1	710	8.5	6,782	46.9	19,694	26.7
NHS Highland	261,627	5.1	4,874	5.4	45	4.1	1,254	15.0	87	0.6	3,791	5.1
NHS Lanarkshire	535,951	10.5	11,040	12.2	254	23.3	2,308	27.6	506	3.5	8,764	11.9
NHS Lothian	755,736	14.8	8,679	9.6	313	28.7	970	11.6	556	3.8	7,328	9.9
NHS Orkney	16,158	0.3	309	0.3	8	0.7	25	0.3	<5	-	280	0.4
NHS Shetland	17,663	0.3	357	0.4	<5	-	17	0.2	90	0.6	284	0.4
NHS Tayside	346,220	6.8	6,050	6.7	85	7.8	1,115	13.3	106	0.7	5,041	6.8
NHS Western Isles	16,898	0.3	391	0.4	<5	-	6	0.1	41	0.3	363	0.5
Unknown	660,586	12.9	3,721	4.1	197	18.0	304	3.6	732	5.1	2,864	3.9

The table presents the number and percentage of individuals in each category indicated by the column headings. Percentages in the 'Total' row reflect the share of individuals in each category as a proportion of the total population. Cell counts <5 have been suppressed.

Cases of long COVID identified by sick notes will be influenced by variation in the share of working age individuals resident in each health board. Due to low usage of long COVID clinical codes, variation across health boards could be influenced by the coding practices of a small number of clinicians.

References

1 Clift A K, Coupland C A C, Keogh R H, Diaz-Ordaz K, Williamson E, Harrison E M et al. Living risk prediction algorithm for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study BMJ 2020; 371 :m3731 doi:10.1136/bmj.m3731.

2 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995 Jan;57(1):289-300.

[3] National Institute for Health Research. Payment Guidance for Researchers and Professionals. Southampton (UK): National Institute for Health Research; 2020. Available from: https://www.nihr.ac.uk/documents/payment-guidance-for-researchers-and-professionals/27392#payment-rates. [Accessed 1 March 2023].

[4] Long Covid Cross-Party Group. Get Involved: Cross Party Groups. 2021 [online] Available at: https://www.parliament.scot/get-involved/cross-party-groups/current-and-previous-cross-party-groups/2021/long-covid-cross-party-group/ [Accessed 1 March 2023].

[5] White C, O'Boyle J, Ormerod J, & Summers R. Hearing our voices - Long Covid: The impact on our lives. 2022. Long COVID Scotland.