BMJ Open

How far back do we need to look to capture diagnoses in electronic health records? A retrospective observational study of hospital electronic health record data

Jadene Lewis,[1,2] Felicity Evison [ID] ,[1,2] Rominique Doal,[1,2] Joanne Field,[3]
Suzy Gallier,[1,2] Steve Harris,[4,5] Peta le Roux,[3] Mohammed Osman,[6,7]
Chris Plummer,[3,7] Elizabeth Sapey [ID] ,[1,8] Mervyn Singer,[4,9] Avan A Sayer,[6,7]
Miles D Witham [ID] ,[6,7] The ADMISSION Research Collaborative

For numbered affiliations see end of article.

Correspondence to
Professor Miles D Witham;
miles.witham@newcastle.ac.uk

ABSTRACT

Objectives Analysis of routinely collected electronic health data is a key tool for long-term condition research and practice for hospitalised patients. This requires accurate and complete ascertainment of a broad range of diagnoses, something not always recorded on an admission document at a single point in time. This study aimed to ascertain how far back in time electronic hospital records need to be interrogated to capture long-term condition diagnoses.

Design Retrospective observational study of routinely collected hospital electronic health record data.

Setting Queen Elizabeth Hospital Birmingham (UK)-linked data held by the PIONEER acute care data hub.

Participants Patients whose first recorded admission for chronic obstructive pulmonary disease (COPD) exacerbation (n=560) or acute stroke (n=2142) was between January and December 2018 and who had a minimum of 10 years of data prior to the index date.

Outcome measures We identified the most common International Classification of Diseases version 10-coded diagnoses received by patients with COPD and acute stroke separately. For each diagnosis, we derived the number of patients with the diagnosis recorded at least once over the full 10-year lookback period, and then compared this with shorter lookback periods from 1 year to 9 years prior to the index admission.

Results Seven of the top 10 most common diagnoses in the COPD dataset reached >90% completeness by 6 years of lookback. Atrial fibrillation and diabetes were >90% coded with 2–3 years of lookback, but hypertension and asthma completeness continued to rise all the way out to 10 years of lookback. For stroke, 4 of the top 10 reached 90% completeness by 5 years of lookback; angina pectoris was >90% coded at 7 years and previous transient ischaemic attack completeness continued to rise out to 10 years of lookback.

Conclusion A 7-year lookback captures most, but not all, common diagnoses. Lookback duration should be tailored to the conditions being studied.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We analysed common conditions associated with two different index conditions underlying hospital admission.
⇒ Our mature, large hospital dataset enabled us to study a variety of conditions and analyse longitudinal trends over 10 years.
⇒ Less common associated conditions were not studied and may require different lookback periods for effective capture.
⇒ Our analysis was confined to a single electronic hospital record and the required lookback periods could differ between hospitals.

BACKGROUND

Multiple long-term conditions (MLTCs) (multimorbidity) are an important but understudied challenge to healthcare systems.[1] MLTCs, defined as two or more long-term health conditions, are common with an estimated global prevalence of 37%.[2] The prevalence in old age is even higher, reaching 80% by the age of 80 years in the UK.[3] Much of the research on MLTCs to date has used large population-based datasets or has focused on primary care data; there has been comparatively little work on MLTCs within the hospital environment.[4] Understanding the associations, consequences and pathways of care for MLTCs within hospital is critical if we are to understand how best to prevent the adverse consequences of MLTCs and deliver better care for hospitalised patients.[5]

Analysis of routinely collected electronic healthcare records (EHRs) affords the opportunity to study MLTCs at scale, across a wide range of patient groups. Use of routine data

enables inclusion in analyses of those who would not be able to join traditional consented studies, such as those with dementia or with critical illness. The accurate identification of which long-term conditions affect hospitalised patients is central to the study of MLTCs in patients admitted to hospital. In the UK, diagnoses within EHRs are recorded at discharge, but can also be recorded during admissions. The quality and completeness of diagnostic coding are highly variable however,[6 7] and examining diagnoses coded during a single admission may not provide a complete picture of all diagnoses affecting an individual.

Aggregating diagnoses from a series of hospital admissions provides an opportunity to improve the completeness of recording, but it is unclear how far back through an EHR researchers should look. Different EHRs have been in existence for different amounts of time, and comparison between EHRs may introduce bias if different lookback times introduce differences in the completeness of diagnostic recording. Understanding the completeness of diagnostic recording for a given lookback period would help to avoid under-recording of diagnoses in both research analyses and clinical applications. This in turn would enable analysis teams to make informed judgements about what records and which conditions to study when analysing routinely collected electronic health data.

The aim of this analysis was therefore to investigate how far back we need to look in order to capture the majority of diagnoses documented from previous admissions, and whether there are differences between diagnoses in what the minimum lookback period should be.

METHODS

Hospital and data sources

The Queen Elizabeth Hospital Birmingham (QEHB) is a National Health Service (NHS) acute hospital providing secondary and tertiary care services to adults in a large urban and suburban catchment area in England, UK. The hospital has 1269 beds including 80 level 2/3 intensive care beds and an emergency department that assesses >300 patients per day. QEHB uses an EHR (PICS, Birmingham Systems) containing time-stamped, structured records including demographics, location, time of admission and discharge. It also contains details of all treatments and investigations, and physiological measurements such as pulse, blood pressure and respiratory rate derived from the National Early Warning Score version 2. Diagnoses are currently coded using the Systematised Nomenclature of Medicine–Clinical Terms (SNOMED) system, and historically used International Classification of Diseases version 10 (ICD-10). The EHR has been in place since 1999 and the Trust is a paperless environment for all care provision and planning.

For this analysis, we used two data extracts. We extracted data on all acute admissions between 1 January 2018 and 31 December 2018 with a discharge diagnosis of chronic obstructive pulmonary disease (COPD) exacerbation (as

a primary or secondary diagnosis). We also used SNOMED codes recorded in the EHR to resolve whether a diagnosis coded as 'COPD unspecified' was due to acute exacerbation. Using this time period enabled us to avoid any distorting effects from the COVID-19 pandemic. We took the date of the first hospital admission for COPD exacerbation in 2018 as the index date, and extracted data on age, sex and all hospital diagnoses coded by ICD-10 discharge codes within 10 years prior to the admission under study. For the second data extract, we repeated the procedure but extracted data on all acute admissions with a discharge diagnosis of stroke. We chose COPD exacerbation and stroke as exemplar conditions as they are common causes of admission to hospital and are associated with a substantial burden of comorbidity.

Approval by PIONEER Data Trust Committee

Governance and linkage processes for PIONEER data have been published previously.[8]

Patient and public involvement

The ADMISSION grant programme was co-designed with a panel of patients including patients with MLTCs. Two public co-applicants form part of the management team for the ADMISSION Programme, and studies within ADMISSION including this analysis are discussed with the ADMISSION Patient Advisory Group, who provide input on the design, interpretation and dissemination of results.

Analyses

We derived diagnoses made during, or at discharge, from ICD-10 codes recorded in the EHR over the 10 years prior to the index admission date. Code lists used for each diagnosis are given in online supplemental material; the conditions selected were based on conditions previously used in UK Biobank,[9] modified to remove diagnoses less likely to be coded in hospital discharge data. Code lists were reviewed by clinicians in the research team prior to data extraction. We then examined the percentage of diagnoses that were captured by looking back from the date of index admission for increasing periods of time from 1 year to 10 years. We compared the proportion of patients with diagnoses captured for each lookback period with the number of diagnoses captured with 10 years of lookback. Only individuals with a full 10 years of EHR data were included in the analyses. We chose a threshold of 90% completeness against which to report results; this threshold was agreed by clinician authors as representing a balance between accuracy and achievability.

RESULTS

We included 560 individuals in the COPD exacerbation analysis dataset (32 were excluded due to incomplete lookback data) and 2142 individuals in the acute stroke dataset (189 were excluded due to incomplete lookback data). A flow chart depicting selection of patient
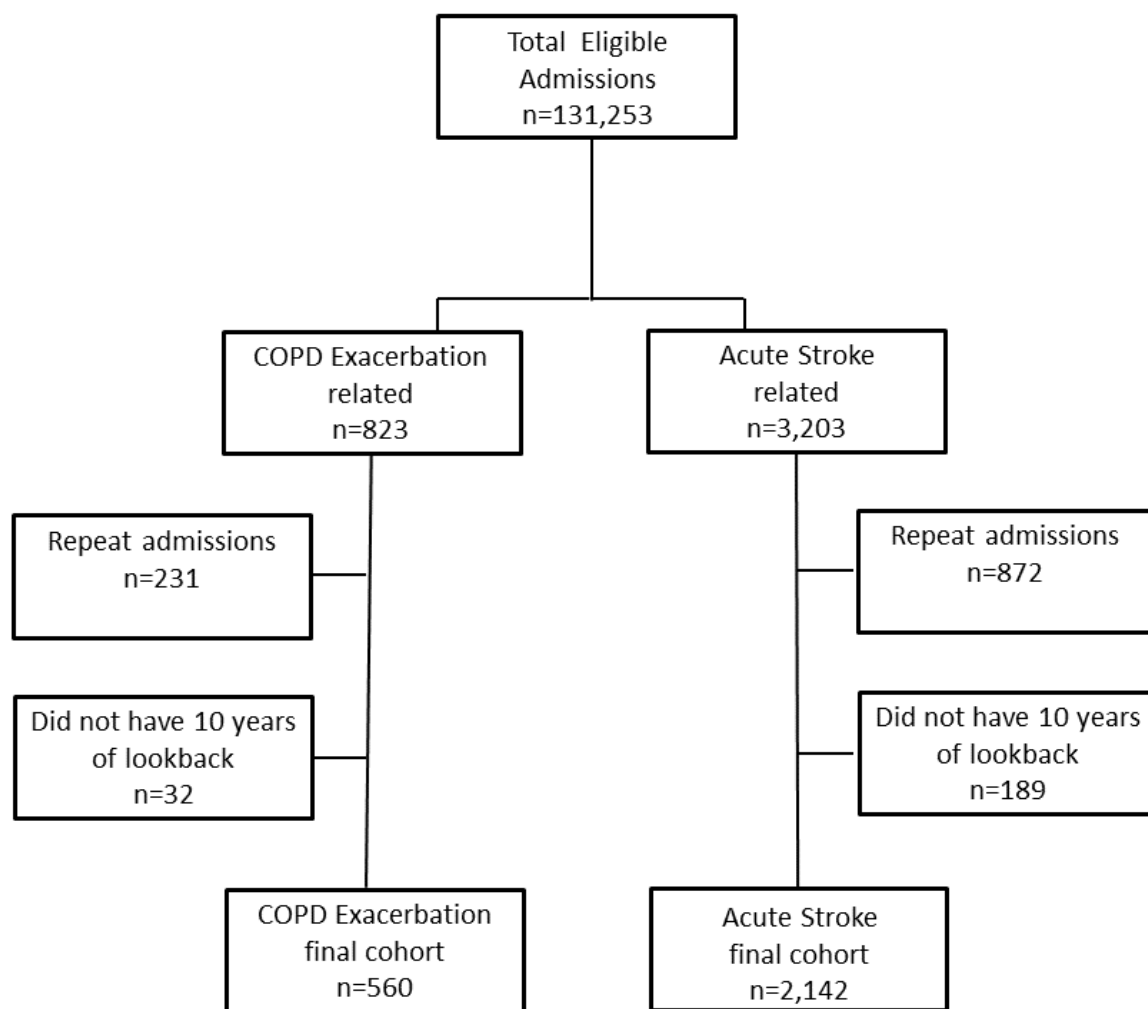
Figure 1 Flow chart showing selection of patient episodes for inclusion in analysis. COPD, chronic obstructive pulmonary disease.

admissions is shown in figure 1. A total of 73 different diagnoses were recorded for those with COPD exacerbation (median three per individual) and a total of 80 different diagnoses were recorded for those with acute stroke (median six per individual). Table 1 shows details of individuals in each dataset, along with the top 10 most common diagnoses recorded in each dataset.

Figure 2 shows the percentage of each condition detected per year of lookback in the COPD dataset, using the number recorded over the full 10-year lookback period as reference. Seven of the top 10 reached 90% completeness by 6 years, but two conditions (asthma and hypertension) continued to show increases in coding completeness with increasing duration of lookback through to 10 years. Figure 3 shows similar data for the acute stroke dataset; there was considerable heterogeneity, from dementia (90% by 3 years) and myocardial infarction (90% by 4 years) through to diabetes mellitus (90% by 5 years) and angina pectoris (90% by 7 years). Previous transient ischaemic attack (TIA) coding completeness continued to increase through to 10 years with no flattening the of curve.

DISCUSSION

Both cohorts in our study had a high prevalence of MLTCs, reinforcing the need for methodological studies in this area to ensure the robustness of analyses using routine data. Our results suggest that a 7-year lookback is sufficient to capture some but not all common diagnoses for patients admitted with COPD exacerbation or acute stroke. For most diagnoses investigated, the prevalence curves flatten out between 6 and 7 years, suggesting that additional years of lookback add few additional diagnoses. There were however notable exceptions, particularly hypertension and asthma in the COPD cohort, and previous TIA in the acute stroke cohort. For these diagnoses, the curves did not flatten at 10 years, suggesting that even longer periods of lookback may be required to gain complete diagnostic data. These data suggest that the duration of lookback needs to be tailored to the condition of interest.

Our results give a useful indication of what the minimum lookback time in hospital EHRs might need to be, but we are not able to ascertain why lookback time varies between diagnoses. Conditions which are the

Table 1 Study population details

	COPD dataset		Stroke dataset	
N	560		2142	
Mean age (years) (SD)	71 (11)		73 (13)	
Female sex (%)	296 (52)		1064 (50)	
Median number of comorbid conditions (IQR)*	3 (1–5)		6 (4–8)	
Top 10 previously diagnosed conditions†	Name	Frequency (%)	Name	Frequency (%)
1	COPD	560 (100)	Stroke	1729 (81)
2	Heart failure	227 (41)	Atrial fibrillation	843 (39)
3	Asthma	206 (37)	TIA	832 (39)
4	Atrial fibrillation	197 (35)	All diabetes	804 (38)
5	Depression	182 (33)	Heart failure	631 (29)
6	Anxiety	180 (32)	COPD	525 (25)
7	All diabetes	159 (28)	Myocardial infarction	519 (24)
8	Hypertension	151 (27)	Angina	491 (23)
9	Myocardial infarction	137 (24)	Dementia	432 (20)
10	Angina	126 (23)	Depression	426 (20)

*In addition to the index condition (COPD or stroke).
†Recorded at an admission prior to the index admission.
COPD, chronic obstructive pulmonary disease; TIA, transient ischaemic attack.

direct cause of hospital admission may be more likely to be recorded, and patients who have frequent hospital admissions may be more likely to have their associated diagnoses recorded. Conditions relevant to a particular specialty (for example, respiratory diagnoses when admitted under a respiratory team) may be more likely to
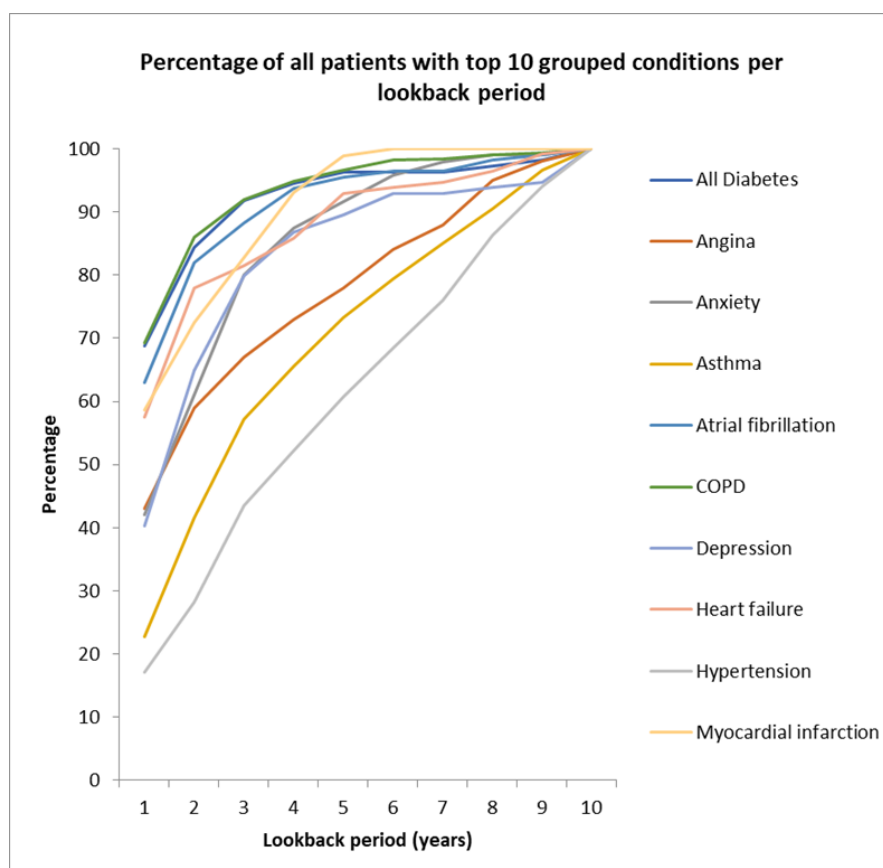
Figure 2 COPD exacerbation top 10 previously diagnosed conditions. COPD, chronic obstructive pulmonary disease.
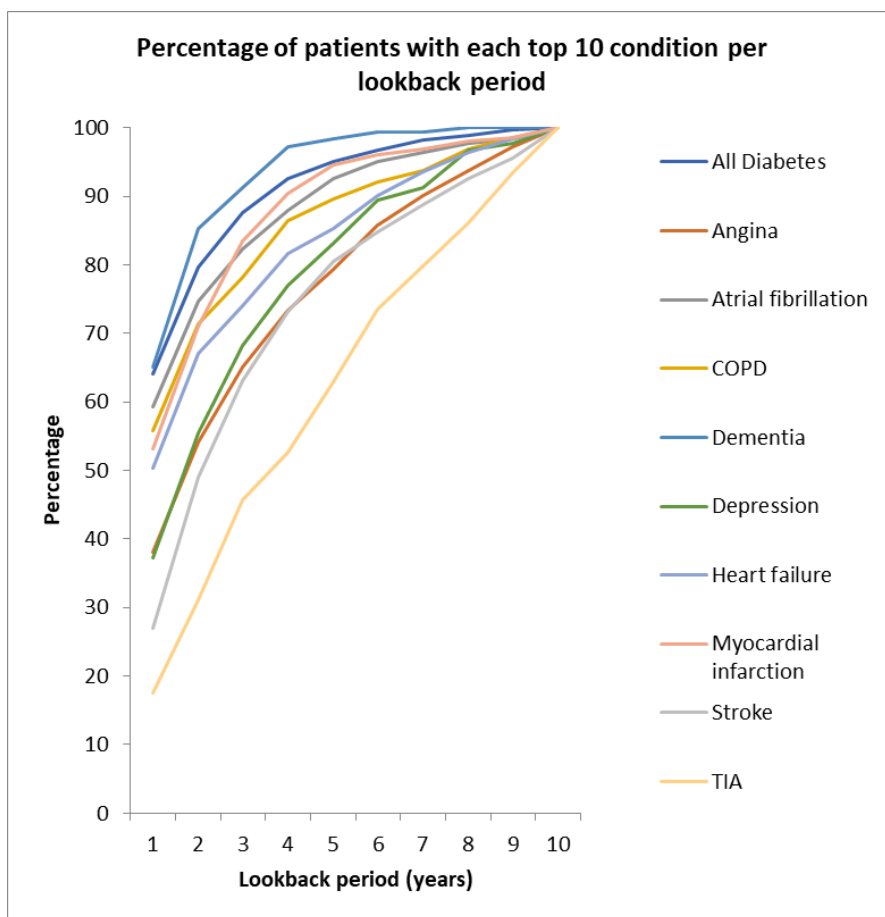
Figure 3 Acute stroke top 10 previously diagnosed conditions. COPD, chronic obstructive pulmonary disease; TIA, transient ischaemic attack.

be recorded than those not associated with the admitting specialty. Further work is required to test these hypotheses. For some diagnoses, reliance on coded, structured diagnostic data is likely to remain inadequate as not all diagnoses are written down in hospital records. Even for those diagnoses that are written down, many are not recorded or coded in a structured manner.[10] Extracting these diagnoses requires additional work, for instance, by interrogating unstructured EHR data using techniques such as natural language processing,[11] by using additional data sources, for instance, prescribing records[12] or laboratory data,[13] and by linkage to alternative sources of diagnostic data (for instance, data from primary care).[14] Our results help to highlight diagnoses for which this work is particularly needed.

Strengths of our approach include the use of routinely collected electronic healthcare data and the interrogation of two common diagnostic reasons for hospital admission. Using routinely collected electronic health data for MLTC research enables inclusion of a wide range of patients. This includes groups excluded from traditional cohort studies because of inability to consent due to illness or cognitive impairment, or who belong to other groups underserved by research such as socioeconomically disadvantaged groups.[15]

Routine data also enable the use of very large sample sizes and can provide detailed longitudinal follow-up incorporating a wide range of conditions and outcomes. We were unable to look back further than 10 years without a reduction in the number of patients included in the analysis, and we have confined our analysis to the most common conditions associated with each index condition. Other less common conditions may require different lookback periods. Our choice of conditions was based on those previously analysed in UK Biobank, but UK Biobank does not use an exhaustive list of conditions and it is possible that other common and important conditions were not included in our analysis. Our choice of a 90% threshold for reporting analyses is to a degree arbitrary; different degrees of completeness may be appropriate for different analyses and only individual research or analysis teams can make a judgement as to what degree of completeness is appropriate for a particular analysis. Finally, our analysis is confined to a single hospital, and lookback times may vary across different healthcare organisations, which may vary in how well they code different diagnoses. Future research to replicate our work in different hospitals and in different index conditions would be useful to increase the generalisability of our findings.

In summary, while a 7-year lookback appears adequate to capture some common diagnoses associated with COPD exacerbation and acute stroke admissions, there is considerable variability between diagnoses. These findings will help both research teams and clinical teams decide what length of lookback they need to employ to ensure that capture of diagnoses is sufficiently complete for the intended purpose, whether this is for research, clinical audit or as part of routine healthcare delivery. For EHRs with only a short lookback window, this may mean that certain diagnoses are not amenable to accurate interrogation. The variability in lookback suggests that researchers need to evaluate diagnoses in the specific EHR system that they use to be sure that diagnosis capture is suitably complete for the conditions they wish to study.

Author affiliations
[1] PIONEER Hub, University of Birmingham, Birmingham, UK
[2] Health Informatics, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[3] Digital Services, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
[4] Critical Care Department, University College London Hospitals NHS Foundation Trust, London, UK
[5] Institute of Health Informatics, University College London, London, UK
[6] AGE Research Group, Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
[7] NIHR Newcastle Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust; Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust and Newcastle University, Newcastle upon Tyne, UK
[8] Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK
[9] Bloomsbury Institute of Intensive Care Medicine, University College London, London, UK

and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs

Felicity Evison http://orcid.org/0000-0002-9378-7548
Elizabeth Sapey http://orcid.org/0000-0003-3454-5482
Miles D Witham http://orcid.org/0000-0002-1967-0990

REFERENCES

1 Whitty CJ. Harveian Oration 2017: Triumphs and challenges in a world shaped by medicine. *Clin Med* 2017;17:537–44.
2 Chowdhury SR, Chandra Das D, Sunna TC, *et al*. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. *eClinicalMedicine* 2023;57:101860.
3 Barnett K, Mercer SW, Norbury M, *et al*. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012;380:37–43.
4 Kudesia P, Salimarouny B, Stanley M, *et al*. The incidence of multimorbidity and patterns in accumulation of chronic conditions: a systematic review. *J Multimorb Comorb* 2021;11:26335565211032880.
5 Owen N, Dew L, Logan S, *et al*. Research policy for people with multiple long-term conditions and their carers. *J Multimorb Comorb* 2022;12:26335565221104407.
6 Ugboma I, Syddall HE, Cox V, *et al*. Coding geriatric syndromes: How good are we? *CME J Geriatr Med* 2008;10:34–6.
7 Wright A, McCoy AB, Hickman T-T, *et al*. Problem list completeness in electronic health records: a multi-site study and assessment of success factors. *Int J Med Inform* 2015;84:784–90.
8 Gallier S, Price G, Pandya H, *et al*. Infrastructure and operating processes of PIONEER, the HDR-UK Data Hub in acute care and the workings of the data trust committee: a protocol paper. *BMJ Health Care Inform* 2021;28:e100294.
9 Gallacher KI, McQueenie R, Nicholl B, *et al*. Risk factors and mortality associated with multimorbidity in people with stroke or transient ischaemic attack: a study of 8,751 UK Biobank participants. *J Comorb* 2018;8:1–8.
10 Todd OM, Burton JK, Dodds RM, *et al*. New horizons in the use of routine data for ageing research. *Age Ageing* 2020;49:716–22.
11 Jackson RG, Patel R, Jayatilleke N, *et al*. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ Open* 2017;7:e012012.
12 Hansen ER, Sagi T, Hose K, *et al*. Assigning diagnosis codes using medication history. *Artif Intell Med* 2022;128:102307.
13 Schulz WL, Young HP, Coppi A, *et al*. Temporal relationship of computed and structured diagnoses in electronic health record data. *BMC Med Inform Decis Mak* 2021;21:61.
14 Herrett E, Shah AD, Boggon R, *et al*. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ* 2013;346:f2350.
15 Witham MD, Cooper R, Missier P, *et al*. Researching multimorbidity in hospital: can we deliver on the promise of health informatics? *Eur Geriatr Med* 2023;14:765–8.

Supplementary Information

ADMISSION Research Collaborative Consortium members:

Avan Aihie Sayer[4,5]

Victoria Bartle[8]

Rachel Cooper[4,5]

Heather J Cordell[9]

Ray Holding[8]

Tom Marshall[10]

Fiona E Matthews[11]

Paolo Missier[12]

Ewan Pearson[13]

Chris Plummer[2,5]

Sian Robinson[4,5]

Elizabeth Sapey[1,6]

Mervyn Singer[7]

Thomas Scharf[11]

James Wason[14]

Miles D Witham[4,5]

1. PIONEER Hub, University of Birmingham, Birmingham, UK and Health Informatics, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

2. Digital Services, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

4. AGE Research Group, Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

5. NIHR Newcastle Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust and Newcastle University, Newcastle upon Tyne, UK

6. Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK.

7. University College London Hospitals NHS Foundation Trust, London, UK and Bloomsbury Institute for Intensive Care Medicine, London, UK

8. Public Co-Investigator, ADMISSION Research Collaborative, Newcastle upon Tyne, UK

9. Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

10. Institute of Applied Health Research, University of Birmingham, Birmingham, UK

11. Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

12. School of Computing, Newcastle University, Newcastle upon Tyne, UK

13. Division of Population Health and Genomics, Ninewells Hospital and School of Medicine, University of Dundee, Dundee, UK

14. Biostatistics Research Group, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

ICD-10 codes used to identify conditions

a. List of comorbid conditions with ICD-10 codes used to identify each condition

Stroke	I61, I63.0, I63.1, I63.2, I63.3, I63.4, I63.5, I63.8, I63.9, I69.1, I69.3, G46.3, G46.4, G46.5, G46.6, G46.7, G46.8, I64, I65, I69.4
Transient Ischaemic Attack	G45.0, G45.1, G45.2, G45.3, G45.4, G45.8, G45.9, G46.0, G46.1, G46.2, I65, I66
Subarachnoid haemorrhage	I60, I69.0
Carpal tunnel syndrome	G56.0
Fibromyalgia	M79.7
Disc problem	M50, M51
Prolapsed disc/slipped disc	M50, M51
Ankylosing spondylitis	M45
Osteoarthritis	M15, M16, M17, M18, M19
Gout	M10, M14.0
Trigeminal neuralgia	G50.0
Disc degeneration	M50, M51
Hypertension	I10, I11, I12, I13, I15
Depression	F32, F33
Asthma	J45, J46
Atrial fibrillation	I48
Myocardial infarction	I252, I21, I22, I23, I24.1
Angina	I20.0 I20.1, I20.8, I20.9
Gastro-oesophageal reflux	K21
Oesophagitis	K20, K21.0, K22.1
Barretts oesophagus	K22.7
Gastric ulcer	K25
Gastritis	K29
Duodenal ulcer	K26
Diabetes mellitus	E10, E11, E12, E13, E14, G59.0, G63.2, H28.0, H35.0, H36.0, M14.2, N08.3, O24.0, O24.1, O24.2, O24.3
Hyperthyroidism/thyrotoxicosis	E05.0, E05.1, E05.2, E05.5, E05.8, E05.9, E06.2
Hypothyroidism/myxoedema	E03.5, E03.8, E03.9
Systemic Lupus Erythematosus	L93, M32
Sjogrens syndrome/sicca syndrome	M35.0
Scleroderma/systemic sclerosis	M34
Rheumatoid arthritis	J99.0, M05, M06
Psoriatic arthropathy	M07.0, M07.1, M07.2, M07.3, M09.0, L40.5
Polymyalgia Rheumatica	M31.5, M35.3
Malabsorption/coeliac disease	K90.0
COPD	J40, J41, J42, J43, J44
Anxiety/panic attacks	F40, F41
Obsessive compulsive disorder	F42
Irritable bowel syndrome	K58
Alcohol dependency	I42.6

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Open

Alcoholic liver disease/alcoholic cirrhosis	K70
Renal/kidney failure	N18.5
Kidney nephropathy	N00.0
Diverticular disease	K38.2, K57
Diverticulitis	K38.2, K57
Peripheral vascular disease	I73.1, I73.8, I73.9, I74.3, I74.4, I74.5
Leg claudication/intermittent claudication	I73.1, I73.8, I73.9, I74.3, I74.4, I74.5
Cardiomyopathy	I42.0, I42.6, I25.5, I42.3, I42.5, I42.7, I42.8, I42.9, I43
Hypertrophic cardiomyopathy	I42.1, I42.2
Heart failure/pulmonary oedema	I11.0. I13.0, I13.2, I50
Prostate problem (not cancer)	N40
Enlarged prostate	N40
Benign prostatic hypertrophy	N40
Glaucoma	H40.1, H40.2, H40.9
Epilepsy	G40, G41
Dementia	F00, F01, F03, F05.1, G30, F02.3
Schizophrenia	F20, F21, F22, F25, F28, F29
Mania/bipolar disorder	F30, F31
Manic depression	F30, F31
Dermatitis	L20, L23, L24, L25, L26, L27, L28, L30.0, L30.1, L30.2, L30.5, L30.8, L30.9,
Psoriasis	L40, M07.0, M07.2, M07.3, M09.0
Crohn's disease	K50
Ulcerative colitis	K51
Migraine	G43
Chronic sinusitis	J32
Anorexia	F50.0, F50.1
Bulimia	F50.2, F50.3
Bronchiectasis	J47, Q33.4
Parkinson's disease	F02.3, G20
Multiple sclerosis	G35
Infective/viral hepatitis	B18
Hepatitis B	B18
Hepatitis C	B18
Hepatitis D	B18
Oesophageal varices	I85, I98.2, I98.3
Liver failure/cirrhosis	K70.1, K70.2, K70.3, K70.4, K71.1, K71.7, K72, K74.0, K74.1, K74.2, K74.4, K74.5, K74.6, K76.2, K76.3, B15.0, B16.0, B19.0
Osteoporosis	M80, M81, M82
Chronic fatigue syndrome	G93.3, F48.0
Endometriosis	N80
Meniere's disease	H81.0
Pernicious anaemia	D51
Polycystic ovary	E28.2

Lifetime diagnosis of cancer	Any C code, D00, D01, D02, D03, D04, D05, D06, D07, D08, D09, D49
Fragility fracture	S72.0, S72.1, S72.2, S52.5, S52.6, M48.5, M49.5
Incontinence	N39.3, N39.4
Delirium	F05

b. ICD-10 codes used to identify admissions for index conditions studied:

COPD exacerbation:

- J441,
- 195951007 (SNOMED),
- 196001008 (SNOMED)

Acute stroke:

- I63[subcodes 01234589],
- I693,
- I61
- I691,
- G46[subcodes 345678],
- I64
- I694
- G45[subcodes 0123489]
- G46[subcodes 012]
- I65
- I66