BMJ Open

Development and analysis protocol for ORCHARD-Stroke: a multimodal cohort study of infection, systemic inflammation and poststroke dementia

Elise T Milosevich ,[1] Jim Davies,[2] Kiruba Nagaratnam,[3] Harjit Bains,[4] Aubretia McColl,[5] Sarah T Pendlebury [6,7]

To cite: Milosevich ET, Davies J, Nagaratnam K, *et al*. Development and analysis protocol for ORCHARD-Stroke: a multimodal cohort study of infection, systemic inflammation and poststroke dementia. *BMJ Open* 2026;**16**:e128641. doi:10.1136/bmjopen-2026-128641

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2026-128641).

Received 20 September 2026
Accepted 05 November 2026

Check for updates

© Author(s) (or their employer(s)) 2026. Re-use permitted under CC BY. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Dr Sarah T Pendlebury; sarah.pendlebury@ndcn.ox.ac.uk

ABSTRACT

Introduction Cognitive impairment and dementia affect approximately one-third of stroke survivors. Established risk factors, including age, stroke severity and pre-existing cerebrovascular disease, account for only part of the variance, and potentially modifiable contributors remain poorly characterised. In non-stroke populations, infection is associated with increased dementia risk, particularly vascular dementia, and risk may be greatest in patients with underlying cerebral small vessel disease (SVD). Infection complicates around one-third of acute strokes and SVD is prevalent in this population. However, few studies have examined infection and poststroke dementia, and the potential modifying effects of acute inflammation and SVD burden are unknown. Oxford and Reading Cognitive Comorbidity, Frailty and Ageing Research Database-Stroke (ORCHARD-Stroke) aims to address these gaps using large-scale multimodal routinely acquired data comprising hospital electronic health records (EHRs) with linked mental health records, Sentinel Stroke National Audit Programme (SSNAP) data and brain imaging, allowing detailed phenotyping of acute stroke at scale together with infection, inflammatory biomarkers and cognitive outcomes.

Methods and analysis ORCHARD-Stroke is an observational cohort study using routinely acquired EHR and audit data. All primary acute stroke admissions will be extracted from the EHRs of Oxford University Hospitals National Health Service Foundation Trust and Royal Berkshire National Health Service Foundation Trust from January 2017 to December 2030 (n≥10 000). Eligible patients will have a primary International Classification of Diseases, 10th Revision diagnosis of ischaemic stroke or intracerebral haemorrhage. Subarachnoid haemorrhage is excluded. EHR data will be linked to SSNAP for stroke severity and subtype, and to regional mental health records for cognitive outcomes. Routinely acquired brain imaging will be obtained with automated quantification of SVD and atrophy, and acute ischaemic lesion extent from existing e-Alberta Stroke Program Early Computed Tomography Score (e-ASPECTS) outputs. The primary exposure is in-hospital infection, classified by site, severity and number. The primary outcome is incident poststroke dementia, after exclusion of prestroke dementia, analysed using Fine and Gray competing risks regression with death as the competing event and adjusted for prespecified covariates including demographics, stroke severity, SVD burden, comorbidity and frailty. Effect modification by SVD burden will be tested using a formal interaction term, and analyses stratified by stroke subtype and by presence or absence of moderate-to-severe SVD. Secondary analyses will examine inflammatory biomarkers (acute phase proteins, blood count indices, composite inflammatory indices) with and without co-occurring infection, and outcomes including cognitive decline and dementia subtype.

Ethics and dissemination ORCHARD-Stroke operates under the Oxford and Reading Cognitive Combordity, Frailty, and Ageing Research Database-Electronic Patient Records

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Oxford and Reading Cognitive Comorbidity, Frailty and Ageing Research Database-Stroke will include a large, unselected cohort of >10 000 patients with acute stroke, with no exclusions based on age, stroke severity or capacity, maximising generalisability to the regional real-world stroke population.
⇒ Routinely acquired hospital electronic health records (EHRs) will be linked to stroke registry data, brain imaging and mental health trust records, providing multimodal data with detailed phenotyping of stroke events at scale: severity and subtype, small vessel disease (SVD) burden and other brain frailty markers, infection, inflammatory biomarkers and long-term dementia outcomes.
⇒ Infection and inflammatory biomarkers will be evaluated within a single analytical framework as potentially modifiable risk factors for poststroke dementia, with SVD prespecified as an effect modifier and competing risks regression accounting for death as a competing event.
⇒ Linkage to the Sentinel Stroke National Audit Programme will use exact matching on patient identifiers within the secure National Health Service trust environment, but some patients may remain unmatched; missing stroke severity data will be handled by multiple imputation with complete case sensitivity analysis.
⇒ Dementia ascertainment relies on linked EHRs rather than prospective in-person assessment and will miss patients who do not come to the attention of health services. Linkage across multiple data sources including mental health trust records mitigates but does not eliminate this limitation.

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

(ORCHARD-EPR) ethics approval granted by the South Central Oxford C Research Ethics Committee (REC reference: 23/SC/0258). Results will be disseminated through peer-reviewed publication and presentations at national and international stroke and dementia conferences.

INTRODUCTION

Poststroke dementia affects approximately one-third of stroke survivors and is a leading cause of acquired disability, institutionalisation and caregiver burden worldwide.[1–3] Risk is mediated in part by pre-existing cerebral small vessel disease (SVD) and other markers of brain frailty and by the direct effects of the stroke lesion, yet these factors account for only a proportion of the observed variance and potentially modifiable contributors remain poorly characterised.[4 5]

Approximately 30% of patients hospitalised with stroke develop infection during their admission.[6] In non-stroke populations, systemic infection is associated with increased dementia risk, with dose-response effects and preferentially greater risk for vascular than Alzheimer's disease subtypes, implicating endothelial activation, blood-brain barrier (BBB) disruption and SVD progression as potential mechanisms.[7–9] Patients with stroke may be particularly susceptible to the cognitive effects of infection because of stroke-induced immunosuppression, frequent co-existent SVD and increased BBB permeability in the acute period.[10 11]

Evidence in stroke populations has been limited to small studies or retrospective administrative data without detailed stroke severity, infection severity or consideration of SVD.[12–15] Our recent systematic review and meta-analysis found consistent associations between poststroke infection and adverse cognitive outcomes, with evidence of dose-response effects by infection severity and number of episodes and greatest susceptibility in patients with SVD. Certain acute inflammatory biomarkers were also associated with cognitive outcomes, but the evidence was limited overall and far fewer studies examined infection than inflammatory biomarkers. No study considered the two exposures together despite their frequent co-occurrence, leaving their relative contribution unresolved.[16] Large-scale studies in stroke cohorts with robust poststroke dementia ascertainment are therefore needed, jointly evaluating infection and inflammatory markers as potentially modifiable exposures, with adjustment for the key confounder of stroke severity and evaluation of SVD as an effect modifier.

We will address these gaps by creating a large multimodal database, the Oxford and Reading Cognitive Comorbidity, Frailty and Ageing Research Database-Stroke (ORCHARD-Stroke), exploiting routinely acquired data on patients with acute stroke presenting to secondary care. Individual patient electronic health records (EHRs) from Oxford University Hospitals and Royal Berkshire NHS Foundation Trusts, serving a combined regional (Oxfordshire and Berkshire) catchment population of approximately 1.8 million, will be linked to Sentinel Stroke National Audit Programme (SSNAP) data, routinely acquired neuroimaging and mental health trust records. ORCHARD-Stroke will therefore include detailed phenotyping of the acute stroke event with infection and inflammation markers, quantification of SVD and atrophy and long-term cognitive outcomes, distinguishing it from large administrative datasets focused on hospital International Classification of Diseases, 10th Revision (ICD-10) coding.

AIMS AND OBJECTIVES

The primary aim of ORCHARD-Stroke is to determine whether acute infection and inflammation, measured using routinely acquired biomarkers, are associated with poststroke dementia independent of prespecified covariates, and whether these associations are modified by SVD burden.

The main objectives are to:

1. determine the association between in-hospital infection and poststroke dementia, including by infection site, severity and number, adjusted for stroke severity, SVD burden, frailty, comorbidity and vascular risk factors, and stratified by stroke subtype;
2. determine whether associations between in-hospital infection and poststroke dementia are modified by the presence of moderate-to-severe SVD;
3. determine the association between acute inflammatory biomarkers and poststroke dementia, with and without co-occurring infection.

We hypothesise that:

1. in-hospital infection will be associated with increased poststroke dementia risk in a dose-response manner, with stronger associations in patients with moderate-to-severe SVD[12–14];
2. routinely acquired acute inflammatory biomarkers will show weaker associations with poststroke dementia than in-hospital infection and will not be independently associated after adjustment for co-occurring infection.[13]

METHODS AND ANALYSIS

Study design and setting

ORCHARD-Stroke includes de-identified EHRs (Cerner Millennium) on individual patients with acute cerebrovascular events admitted to the regional acute stroke units or other acute wards at the Oxford University Hospitals NHS Foundation Trust (OUHFT) and Royal Berkshire NHS Foundation Trust (RBFT) (figure 1). Data are obtained by the hospital information analysts, identifiers are removed and patients are assigned a study ID prior to transfer to the database. Extracted data are checked for missing values and sense-checked by STP in collaboration with the hospital information analysts, as established for Oxford and Reading Cognitive Combordity, Frailty, and Ageing Research Database-Electronic Patient Records (ORCHARD-EPR).[17] Data extraction covers the period

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

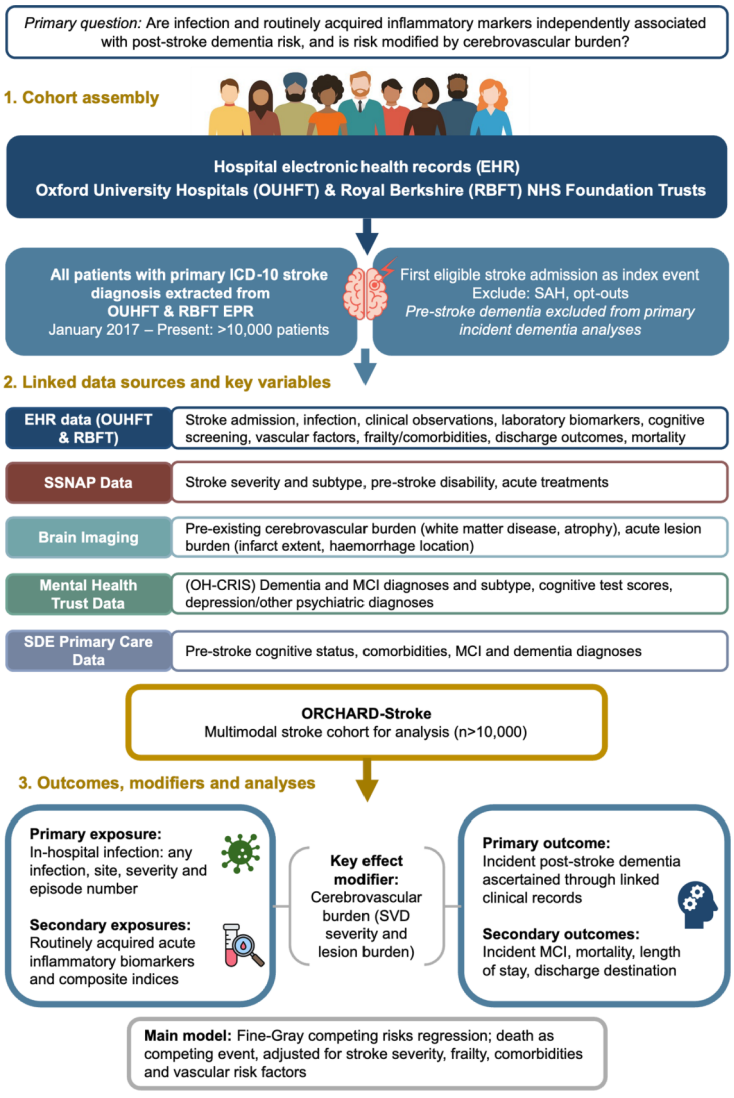


Figure 1 ORCHARD-Stroke study design, data linkage and analytical framework. Linkage to SDE primary care data is planned from 2027/28 and is not required for the analyses described in this protocol. The secondary outcome shown as incident MCI comprises incident cognitive impairment, defined as MCI, dementia or a low cognitive test score (table 1). EHR, electronic health record; EPR, electronic patient record; ICD-10, International Classification of Diseases, 10th Revision; MCI, mild cognitive impairment; OH-CRIS, Oxford Health Clinical Record Interactive Search; ORCHARD-Stroke, Oxford and Reading Cognitive Comorbidity, Frailty and Ageing Research Database-Stroke; OUHFT, Oxford University Hospitals NHS Foundation Trust; RBFT, Royal Berkshire NHS Foundation Trust; SAH, subarachnoid haemorrhage; SDE, secure data environment; SSNAP, Sentinel Stroke National Audit Programme; SVD, small vessel disease.

from 1 January 2017, when EHRs were fully embedded to 31 December 2030, with data refreshed annually. Ethics approval was obtained from the South Central Oxford C Research Ethics Committee (REC reference: 23/SC/0258). Informed consent was not required but the study was advertised on the hospital websites and via posters and patients could opt out.

Participants and eligibility

All adult patients admitted to OUHFT or RBFT between January 2017 and the most recent available data extraction are eligible if they have a primary ICD-10 diagnosis of ischaemic stroke or intracerebral haemorrhage (table 1). The core ORCHARD-EPR extract comprises all patients aged ≥65 years together with younger patients who have a completed cognitive screen[17]; a supplementary extract of stroke admissions in patients aged <65 years will be requested from the hospital information analysts so that ORCHARD-Stroke includes all adult stroke admissions without age restriction. Subarachnoid haemorrhage is excluded because of its distinct pathophysiology, clinical course and cognitive trajectory. For patients with more than one stroke admission during the study period, the first admission will be used as the index event for primary analyses. Subsequent stroke admissions will be retained

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 Variables, definitions and data sources

Variable	Operational definition	Data source	Timing
Exposures			
In-hospital infection (primary exposure)	Any infection recorded during the index admission, identified using predefined ICD-10 diagnostic code lists and supplemented by SSNAP infection variables, including antibiotic prescribing during admission	EHR; SSNAP	Index admission
Infection site	Hierarchical classification as respiratory (including pneumonia), urinary tract, bloodstream, skin and soft tissue, gastrointestinal, central nervous system or other, adapted from previously published classifications[7 13]	EHR	Index admission
Infection severity	SIRS criteria[26] derived algorithmically from serial physiological observations and laboratory results, and NEWS[27]	EPR observations and laboratory results	Serial, index admission
Number of infection episodes	Count of distinct infection episodes during the index admission, analysed as an ordinal variable	EHR	Index admission
Postdischarge infection	Hospital admission with a recorded infection diagnosis, identified using the same predefined diagnostic criteria	EHR	≥3 months after index admission
Acute phase proteins	C-reactive protein, fibrinogen and albumin concentrations	EHR laboratory results	Admission; peak; day 3
Blood count indices	White cell, neutrophil, lymphocyte, monocyte and platelet counts	EHR laboratory results	Admission; peak; day 3
Composite inflammatory indices	NLR and SII (calculated as neutrophil count multiplied by platelet count divided by lymphocyte count), derived from routine full blood count results	Derived	Admission; peak; day 3
Outcomes			
Incident poststroke dementia (primary outcome)	First recorded dementia diagnosis in any linked data source, identified using predefined ICD-10 diagnostic code lists and recorded mental health service diagnoses; earliest date across sources used	OH-CRIS; EHR	Discharge to censoring
Incident cognitive impairment	Composite of mild cognitive impairment, dementia or a low cognitive test score on routine cognitive screening	OH-CRIS; EHR	Discharge to censoring
All-cause mortality	Date of death	Office for National Statistics	To censoring
Length of stay; discharge destination	Duration of index admission in days; discharge destination category	EHR	Index admission
Longitudinal cognition	Serial AMTS recorded at subsequent hospital admissions	EHR	After incident cognitive impairment
Covariates			
Demographics	Age and sex	EHR	Index admission
Socioeconomic deprivation	English IMD decile derived from residential postcode	EHR	Index admission
Stroke severity	NIHSS score	SSNAP	Index admission and discharge
Stroke subtype	Modified TOAST criteria where available,[28] supplemented by routinely recorded stroke diagnoses	SSNAP; EHR	Index admission
Small vessel disease burden	Fazekas or ARWMC score, or automated CT-derived quantification	Brain imaging	Index admission or preceding 12 months
Brain atrophy	GCA score or automated CT-derived quantification	Brain imaging	Index admission or preceding 12 months
Acute lesion extent	ASPECTS[29] obtained using Brainomix e-ASPECTS[24]	Brain imaging	Index admission

Continued

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 Continued

Variable	Operational definition	Data source	Timing
Prestroke frailty	Clinical Frailty Scale[30] and Hospital Frailty Risk Score, a validated ICD-10 coding algorithm[31 32]	EHR	Pre-admission
Comorbidity	Charlson Comorbidity Index derived using the validated ICD-10 coding algorithm with updated weights[33]	EHR	Pre-admission
Vascular risk factors	History of hypertension, diabetes mellitus, atrial fibrillation, hyperlipidaemia, peripheral arterial disease or previous stroke or transient ischaemic attack, identified using predefined ICD-10 diagnostic code lists and SSNAP fields	EHR; SSNAP	Pre-admission
Smoking status	Current, former or never smoker, from routinely recorded clinical data and SSNAP fields	EPR; SSNAP	Pre-admission
Prestroke function	mRS	SSNAP	Pre-admission
Dysphasia at onset	Presence or absence of dysphasia recorded at presentation	SSNAP	Index admission
Care home residence	Care home residence prior to admission	EHR	Pre-admission
Prestroke and acute cognition	AMTS from routine cognitive screening, recorded dementia diagnosis and delirium diagnosis during admission	EHR; OH-CRIS	Pre-admission; index admission

AMTS, Abbreviated Mental Test Score; ARWMC, age-related white matter changes; ASPECTS, Alberta Stroke Program Early CT Score; EHR, electronic health record; EPR, electronic patient record; GCA, global cortical atrophy; ICD-10, International Classification of Disease,10th Revision; IMD, index of multiple deprivation; mRS, modified Rankin Scale; NEWS, National Early Warning Score; NIHSS, National Institutes of Health Stroke Scale; NLR, neutrophil-to-lymphocyte ratio; OH-CRIS, Oxford Health Clinical Record Interactive Search; SII, Systemic Immune-Inflammation Index; SIRS, systemic inflammatory response syndrome; SSNAP, Sentinel Stroke National Audit Programme; TOAST, Trial of Org 10172 in Acute Stroke Treatment.

to define recurrent stroke during follow-up and for sensitivity analyses.

Dataset assembly and data linkage

Linkage to SSNAP

Both OUHFT and RBFT submit patient-level data to SSNAP as part of mandatory national audit participation, and each trust holds a local extract of its own submitted records (figure 1). Linkage between the trust SSNAP extract and ORCHARD-EPR will be performed by exact (deterministic) matching on hospital number, with date of birth and admission date used to confirm matches, all of which are held in both datasets. Linkage will be undertaken within the secure OUHFT and RBFT NHS environments, with STP the named data controller. Following matching, direct identifiers will be removed, the linked record will be assigned the study ID and only the pseudonymised dataset will be transferred to University of Oxford servers, consistent with the process established for ORCHARD-EPR.[17] Records that cannot be matched will be retained with SSNAP fields missing and vice versa.

SSNAP provides stroke-specific variables not routinely available in hospital EHR data, including National Institutes of Health Stroke Scale (NIHSS) score, stroke subtype, prestroke modified Rankin Scale, acute stroke treatments and stroke unit care metrics. Both OUHFT and RBFT achieve Band A case ascertainment in SSNAP, defined as capture of at least 90% of expected eligible stroke admissions, resulting in near-complete information on stroke severity and subtype in ORCHARD-Stroke.

Linkage to brain imaging

Routinely acquired CT and MRI brain imaging from the index stroke admission will be obtained to quantify pre-existing SVD burden and acute stroke lesion volume as potential effect modifiers and confounders. At OUHFT, imaging is transferred via an automated de-identification pipeline from the Picture Archiving and Communication System to University of Oxford servers.[17] An equivalent process will be used to obtain RBHFT imaging. Where index admission imaging is unavailable or unsuitable for assessment of pre-existing SVD burden, relevant brain imaging from the preceding 12 months will be used.

SVD burden and other markers of brain frailty will be quantified using an automated CT-brain analysis tool developed within the ORCHARD programme, or standard commercial tools for MRI scans. The CT tool quantifies global and temporal lobe atrophy and white matter disease from routinely acquired scans and has been validated against trained human raters in older patients, including an acute stroke cohort.[18] Visual ratings will be used in a subset of scans: the Fazekas or Age-Related White Matter Changes scales for SVD and the Global Cortical Atrophy scale for atrophy to validate the automated outputs.[19–21] Lobar versus deep haemorrhage will be recorded to classify possible cerebral amyloid

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

angiopathy.[22] Acute ischaemic lesion extent will be quantified from existing Brainomix e-Alberta Stroke Program Early Computed Tomography Score (e-ASPECTS) outputs, including the Alberta Stroke Program Early CT Score as a measure of infarct extent where available.[23 24]

Linkage to mental health trust records

Dementia and cognitive impairment outcomes will be ascertained from linked mental health records. For OUHFT patients, linkage is via Oxford Health Clinical Records Interactive Search (OH-CRIS), powered by Akrivia Health under an existing data sharing agreement.[17] OH-CRIS contains de-identified data from inpatient and outpatient mental health records, including dementia and mild cognitive impairment (MCI) diagnoses, cognitive test scores, memory clinic information, depression and other psychiatric diagnoses, derived in part from natural language processing of free-text clinical notes. These data enhance dementia ascertainment beyond ICD-10 hospital coding alone, which is known to be insensitive for early dementia and MCI and provide information on dementia subtype.[25] Equivalent linkage to Berkshire Healthcare NHS Foundation Trust (BHFT) electronic mental health records will be sought under a separate data sharing agreement, to be negotiated independently of the existing OH-CRIS agreement. BHFT is not part of the Akrivia Health CRIS network, its records comprise structured diagnostic and coded data only, without free-text search, so dementia ascertainment for RBFT-linked patients may be less sensitive than for OUHFT-linked patients. Diagnoses recorded before or at the index stroke date will be used to identify prestroke dementia.

Linkage to primary care records

Linkage to primary care data is planned via the regional Secure Data Environment from 2027/28 but is not yet available. Primary care records will provide enhanced information on prestroke and poststroke dementia status, capturing patients who do not have subsequent contact with secondary care, life course exposure to vascular risk factors and comorbidities and infection and antibiotic prescribing on follow-up. These data are not required for the analyses described in this protocol.

Variables

Exposures: infection

The primary exposure is in-hospital infection during the index stroke admission, identified using routinely collected clinical and diagnostic data as detailed in table 1.[26–35] Infection site will be classified using a hierarchical framework adapted from previous work.[7 13] Detailed coding algorithms will be specified in the study analysis plan.

Infection severity will be characterised using routinely recorded physiological and laboratory data. Systemic inflammatory response syndrome (SIRS) criteria will be derived algorithmically from serial physiological observations and laboratory results in EHRs where data completeness permits,[26] and National Early Warning Score (NEWS) will be derived from serial observations recorded during admission.[27] The number of distinct infection episodes during the index admission will be recorded to enable examination of dose-response effects. Any in-hospital infection versus none will be used as a binary primary exposure in the main analysis, with site, severity and episode number examined in secondary analyses.

Postdischarge infection is a prespecified secondary exposure, ascertained from hospital admissions with an infection diagnosis during follow-up. To minimise reverse causation, whereby early cognitive decline after stroke may increase susceptibility to infection, analyses will be restricted to episodes occurring at least 3 months after the index stroke admission, consistent with Morton *et al*.[12]

Exposures: inflammatory biomarkers

Secondary exposures are routinely acquired acute inflammatory biomarkers extracted from EHR laboratory results, comprising acute phase proteins, complete blood count indices and composite inflammatory indices (table 1). Admission values will be used as the primary biomarker time point. Where serial measurements are available, peak values and change from admission to day 3 will be examined in exploratory analyses.

Covariates

Prespecified covariates, with their operational definitions and data sources, are listed in table 1. They comprise demographic factors, socioeconomic deprivation, stroke severity and subtype, brain imaging measures, prestroke frailty and function, comorbidity burden, vascular risk factors, smoking status, care home residence and prestroke and acute cognitive status.

Outcomes

Patients with dementia diagnosed before the index stroke admission are excluded from the primary analyses. The primary outcome is time to incident all-cause poststroke dementia, defined as the interval from discharge after the index stroke admission to the date of first dementia diagnosis recorded in any linked data source: OH-CRIS mental health trust records or a subsequent hospital admission with a dementia ICD-10 code (table 1). The earliest recorded dementia date across all sources will be used.

Two prespecified sensitivity analyses will exclude dementia diagnosed within 3 months of stroke admission, to reduce misclassification of delirium or transient poststroke cognitive change as incident dementia,[12] and within 1 year of stroke admission, to further reduce the possibility that early cognitive decline or preclinical dementia influenced infection risk or dementia ascertainment.[14] Where sample size permits, dementia subtypes will be examined as secondary outcomes given differential

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

infection-dementia risk relationships reported in non-stroke populations.[7 8]

Secondary outcomes are time to incident cognitive impairment, all-cause mortality, length of index hospital stay and discharge destination. Incident cognitive impairment comprises MCI, dementia or a low cognitive test score, ascertained from EHRs and OH-CRIS/BHFT mental health records. Mortality data are obtained from the Office for National Statistics. Longitudinal cognitive trajectory from serial Abbreviated Mental Test Score recordings at subsequent hospital admissions is a prespecified exploratory outcome, recognising that the irregular and non-random timing of readmissions may introduce selection bias.

Statistical analyses

Baseline characteristics will be summarised using descriptive statistics: means with SD or medians with IQRs for continuous variables and frequencies with percentages for categorical variables. Characteristics will be presented overall and according to in-hospital infection status and incident poststroke dementia status. Between-group comparisons will use χ^2 tests for categorical variables and t-tests or Mann-Whitney U tests for continuous variables, as appropriate. Patterns of missingness will be reported by trust, calendar year, SSNAP linkage status, imaging availability and outcome linkage source. A flow diagram will be presented reporting the number of patients at each stage of cohort assembly and linkage, including numbers eligible, linked to SSNAP, with available brain imaging, and with linked mental health records.

The primary analysis will use competing risks regression according to the method of Fine and Gray, with death before dementia diagnosis treated as the competing event, to estimate the association between in-hospital infection and time to incident post-stroke dementia.[36] This approach is used in preference to standard Cox regression because treating death as non-informative censoring may overestimate dementia risk in groups with higher mortality, including patients with infection. Follow-up will start at discharge after the index stroke admission because in-hospital infection is defined during the index admission. Patients will be censored at last known contact across linked data sources or at the end of the data extraction window. The primary model will be adjusted for the prespecified covariate set entered simultaneously, and subdistribution HRs with 95% CIs will be reported. Standard Cox proportional hazards regression will be conducted as a sensitivity analysis to facilitate comparison with published literature.

Dose-response analyses will examine whether dementia risk increases with infection severity, defined by SIRS criteria and NEWS, and with the number of infection episodes during the index admission, using episode count as an ordinal variable with a test for linear trend.

NIHSS is available only for SSNAP-linked patients. Missing NIHSS values will be handled using multiple imputation by chained equations, incorporating auxiliary variables expected to predict stroke severity and/or missingness, including admission NEWS, length of stay, discharge destination, stroke ICD-10 subtype and imaging-derived infarct volume where available. A complete case sensitivity analysis restricted to SSNAP-linked patients is prespecified.

The prespecified effect modifier is the presence of moderate-to-severe SVD on brain imaging. A formal interaction term between in-hospital infection and SVD burden will be included in the fully adjusted Fine and Gray model. SVD burden will be dichotomised as moderate-to-severe versus none-to-mild using Fazekas or Age-Related White Matter Changes scale scores, consistent with Pendlebury *et al*.[14] Stratified analyses presenting subdistribution HRs for infection within the group with versus without moderate-to-severe SVD will be used to further explore effect modification. Additional prespecified subgroup analyses will examine infection-dementia associations by stroke subtype: ischaemic versus haemorrhagic, across ischaemic stroke subtypes and by lobar versus deep haemorrhage (cerebral amyloid angiopathy vs hypertensive haemorrhage). Lastly, because dementia ascertainment depth differs by site (OH-CRIS free-text search at OUHFT vs structured-data-only BHFT linkage), a prespecified site-stratified sensitivity analysis will estimate the primary infection-dementia association separately by trust.

Secondary exposure analyses will examine acute inflammatory biomarkers individually and by grouped category using the same Fine and Gray framework. Models including both in-hospital infection and inflammatory biomarkers will test whether biomarker associations are independent of co-occurring infection. Biomarker analyses will also be stratified by the presence versus absence of infection.

Secondary outcome analyses will examine infection and time to incident cognitive impairment using the same competing risks framework. All-cause mortality will be analysed using Cox proportional hazards regression. Length of index hospital stay will be compared by infection status using quantile (median) regression to account for its skewed distribution and discharge destination (home vs other) using logistic regression. Longitudinal cognitive trajectory from serial Abbreviated Mental Test Score recordings will be examined using linear mixed-effects models with random intercepts and slopes. All models will be adjusted for the prespecified covariate set.

As a prespecified exploratory analysis, a poststroke dementia risk prediction model will be developed using penalised regression (Least Absolute Shrinkage and Selection Operator), which shrinks the coefficients of weakly contributing variables towards zero and so selects the most informative predictors from the full covariate set while reducing overfitting. Model discrimination and calibration will be compared with the conventional multivariable model.

All analyses will be conducted in R. Results will be reported in accordance with the Strengthening the

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Reporting of Observational Studies in Epidemiology (STROBE) and the Reporting of Studies Conducted Using Observational Routinely Collected Health Data (RECORD) statements and, for the exploratory prediction model, the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD)-artificial intelligence (TRIPOD+AI) guideline.[37–39]

Sample size

ORCHARD-Stroke uses data acquired in real time during routine clinical care. The sample size is therefore fixed by the size of the source population rather than determined by prospective recruitment, and a formal prospective sample size calculation is not applicable.

The pooled prevalence of prestroke dementia in hospital-based cohorts is 14.4% (95% CI 12.0 to 16.8).[2] After exclusion of prestroke dementia, approximately 8500 of the projected 10 000 patients are expected to be eligible for the primary analysis. Reported poststroke dementia incidence of approximately 10% after first-ever stroke, rising to around one-third after recurrent stroke,[2] indicates in excess of 850 incident dementia events. Effective sample size will nonetheless vary by linkage completeness, missingness and outcome frequency.[40]

As an indicative estimate, assuming 80% power, two-sided alpha of 0.05 and a HR of 1.84 based on Morton *et al*,[12] approximately 5000 patients would be required to detect a comparable effect, well within the projected cohort. This calculation is based on a standard proportional hazards model and is indicative only, since the primary analysis will use Fine and Gray competing risks regression, for which power depends additionally on the rate of the competing event.

Patient and public involvement

The ORCHARD patient and public involvement (PPI) group was established in 2015 and has informed the ORCHARD research programme since its inception.[17] The group includes patients and carers with lived experience of acute hospital admission, stroke and cognitive impairment, alongside a retired general practitioner, Alzheimer's Society representatives and a PPI lead with experience in frailty and dementia studies. PPI discussions have identified cognitive outcomes, dementia risk and prognosis after acute illness and stroke as priorities for patients and families, informing the focus of ORCHARD-Stroke on poststroke dementia. As this is a database study using routinely acquired clinical data without direct patient recruitment, assessment of participant burden is not applicable. The PPI group will advise on dissemination, including lay summaries for local community events and stroke and dementia interest groups.

Ethics and dissemination

ORCHARD-Stroke operates under the existing ORCHARD-EPR ethics approval granted by the South Central Oxford C Research Ethics Committee (REC reference: 23/SC/0258) and the opt-out framework described above. This approval covers disease-specific substudies and linkage of ORCHARD-EPR to additional routinely acquired datasets, including the SSNAP linkage described here.[17] Individual patient consent is not required for inclusion of pseudonymised routinely acquired clinical data, but patients who have opted out are excluded. All data will be processed and analysed within the approved ORCHARD-EPR governance framework, with linkage performed within the secure trust environment and only pseudonymised data transferred to the University of Oxford. Results will be disseminated through peer-reviewed publication, presentations at national and international stroke and dementia conferences, and lay summaries shared with patient, public and charity interest groups, including Alzheimer's Research UK.

Detailed coding algorithms, including predefined ICD-10 code lists where applicable, will be specified in the study analysis plan and are available from the authors on reasonable request.

DISCUSSION

Summary

ORCHARD-Stroke will assemble a large multimodal database to examine infection and acute inflammatory biomarkers as risk factors for poststroke dementia, assessing both exposures within the same analytical framework. By jointly evaluating clinically recorded infection, routinely acquired inflammatory biomarkers, stroke severity and SVD burden, the study addresses key limitations in the existing evidence. Prespecified analysis of effect modification will test whether SVD burden alters infection-associated dementia risk in an unselected real-world stroke population.

Infection as a modifiable risk factor for poststroke dementia

Existing evidence suggests that poststroke infection may contribute to dementia risk, but key methodological limitations remain. Pendlebury *et al*[14] found that infection only predicted dementia in patients with moderate-to-severe SVD, with an adjusted HR of 1.75 (95% CI 1.04 to 2.94; p for interaction=0.023), but the analysis was restricted to transient ischaemic attack and minor stroke. In the largest existing study of unselected patients with stroke, Morton *et al*[12] reported a 44%–84% increase in early poststroke dementia risk after common infections; however, stroke severity and brain imaging markers were unavailable, and infection and dementia were ascertained from primary care coding alone. We previously showed that infection was independently associated with poststroke cognitive impairment at 2 weeks and 6 months, with the strongest associations observed acutely for pneumonia and severe infection, although these attenuated at follow-up.[13] ORCHARD-Stroke

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

will seek to validate and extend these findings to a large, unselected acute stroke population with stroke severity measures, brain imaging markers and long-term dementia ascertainment.

SVD burden as an effect modifier

Underlying SVD may increase susceptibility to the cognitive impacts of infection. SVD is associated with increased BBB permeability, which may facilitate entry of peripheral inflammatory mediators into the brain and accelerate microvascular injury or neurodegenerative change.[9 11 41] The finding that infection predicted dementia only in patients with moderate-to-severe SVD supports this hypothesis but remains untested across the full stroke severity spectrum.[14] If SVD modifies infection-dementia associations after acute stroke, a high-risk subgroup could be identified using routinely acquired CT or MRI brain imaging, with implications for infection prevention, risk stratification and trial recruitment. ORCHARD-Stroke will test this hypothesis across the full stroke severity spectrum and by stroke subtype.

Infection versus inflammatory biomarkers: an unresolved question

Whether infection-associated dementia is driven by the infectious event itself, the accompanying systemic inflammatory response or a combination of both remains unresolved. In our preliminary study of acute stroke, routinely acquired inflammatory biomarkers including C-reactive protein, white cell count and neutrophil-to-lymphocyte ratio were not associated with poststroke cognitive impairment, whereas infection was.[13] Conversely, our systematic review of inflammatory biomarkers and poststroke cognitive outcomes found limited but broadly consistent evidence for an association, but no study has previously compared biomarker associations with and without co-occurring infection.[16] ORCHARD-Stroke will enable assessment of whether routinely acquired inflammatory biomarkers are associated with poststroke dementia independently of infection and will inform future mechanistic studies.

Strengths and limitations

The principal strengths of ORCHARD-Stroke are its size, inclusivity and multimodal data linkage. With >10 000 unselected patients with acute stroke, the cohort will provide statistical power for prespecified analyses by stroke subtype, SVD burden and dementia subtype. The absence of exclusions based on stroke severity, age or decision-making capacity will improve generalisability and reduce the selection biases inherent in consent-based prospective cohorts. Linkage to SSNAP, routinely acquired brain imaging and mental health trust records will allow adjustment for stroke severity, quantification of SVD burden and more accurate dementia ascertainment than hospital coding alone. The prespecification of SVD burden as an effect modifier, the joint analytical framework for infection and inflammatory biomarkers and competing risks regression with death treated as a competing event address key methodological gaps identified in the existing literature.

Several limitations should be acknowledged. Infection ascertainment relies primarily on routinely recorded clinical and diagnostic data rather than microbiological sampling, which may underestimate infection prevalence or misclassify infection site. However, this reflects real-world clinical recognition of infection. In addition, detailed medication data are not yet available beyond the antibiotic use recorded in SSNAP, limiting the ability to adjust for antibiotic treatment. Some patients may not be matched to SSNAP and missing stroke severity data will be handled using multiple imputation. Dementia ascertainment relies on linked clinical records rather than standardised research assessment; however, delirium and cognitive screening in ORCHARD-EPR has been validated against prospectively assessed reference cohorts,[34] and linkage across hospital and mental health trust records will mitigate, although cannot eliminate, missed or delayed diagnoses. Dementia ascertainment may also differ by site, as planned BHFT linkage will provide structured data only, without OH-CRIS's free-text search, although this is addressed with a prespecified site-stratified sensitivity analysis. Future linkage to primary care records will further improve case capture, particularly for patients without subsequent secondary care contact. Education level is not routinely recorded and cannot be included as a covariate, although adjustment for deprivation and baseline cognitive score will partially address this. Smoking status may be underestimated if derived from diagnostic coding or incompletely recorded clinical fields. Apolipoprotein E genotype data are unavailable.

In summary, ORCHARD-Stroke will assemble a large, unselected cohort of patients with acute stroke, linking routinely acquired hospital records to stroke registry data, brain imaging and mental health records to examine infection and routinely acquired inflammatory biomarkers as potentially modifiable risk factors for poststroke dementia. Findings will inform whether those at highest risk can be identified using data already collected in routine care, with implications for infection prevention, risk stratification and future intervention trials.

Author affiliations
[1]Radcliffe Department of Medicine, University of Oxford, Oxford, UK
[2]Department of Computer Science, University of Oxford, Oxford, UK
[3]Stroke Medicine, Royal Berkshire NHS Foundation Trust, Reading, UK
[4]Stroke Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[5]Departments of Acute and Elderly Care Medicine, Royal Berkshire NHS Foundation Trust, Reading, UK

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

[6]Wolfson Centre for Prevention of Stroke and Dementia, University of Oxford, Oxford, UK
[7]Departments of Acute General (Internal) Medicine and Geratology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Contributors ETM and STP jointly conceived the study and designed the protocol. ETM drafted the manuscript. STP critically revised the manuscript, conceived and established ORCHARD-EPR and will supervise the study. JD, KN, HB and AM are project collaborators who reviewed the protocol and provided critical comments. All authors read and approved the final manuscript. STP is the guarantor.

Funding This work was supported by a Race Against Dementia–Alzheimer's Research UK Fellowship awarded to ETM (ARUKRADF2023B-007). STP is supported by the NIHR Oxford Biomedical Research Centre. The assembly of ORCHARD-EPR and its use in the development of dementia risk prediction algorithms for older hospital patients are funded through STP's 5-year NIHR Invention for Innovation (i4i) Programme grant (NIHR204290, 2023–2028).

Disclaimer The views expressed are those of the authors and do not necessarily reflect those of the National Health Service, the National Institute for Health and Care Research or the Department of Health.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the 'Methods' section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Currently, the ethics approval for this study does not permit access to the data by researchers outside the University of Oxford, Oxford University Hospitals NHS Foundation Trust and Royal Berkshire NHS Foundation Trust, although widening access to researchers from other institutions is currently being explored. Predefined lists of diagnostic codes from the International Classification of Diseases, 10th Revision, are specified in the study analysis plan and are available from the authors on request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs
Elise T Milosevich https://orcid.org/0000-0001-9175-0779
Sarah T Pendlebury https://orcid.org/0000-0003-3603-8388

REFERENCES

1 Feigin VL, Brainin M, Norrving B, *et al*. World Stroke Organization: Global Stroke Fact Sheet 2025. *Int J Stroke* 2025;20:132–44.
2 Pendlebury ST, Rothwell PM. Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. *Lancet Neurol* 2009;8:1006–18.
3 Pendlebury ST, Rothwell PM, Oxford Vascular Study. Incidence and prevalence of dementia associated with transient ischaemic attack and stroke: analysis of the population-based Oxford Vascular Study. *Lancet Neurol* 2019;18:248–58.
4 Filler J, Georgakis MK, Dichgans M. Risk factors for cognitive impairment and dementia after stroke: a systematic review and meta-analysis. *Lancet Healthy Longev* 2024;5:e31–44.
5 Rost NS, Brodtmann A, Pase MP, *et al*. Post-Stroke Cognitive Impairment and Dementia. *Circ Res* 2022;130:1252–71.
6 Westendorp WF, Nederkoorn PJ, Vermeij J-D, *et al*. Post-stroke infection: a systematic review and meta-analysis. *BMC Neurol* 2011;11:110.
7 Sipilä PN, Heikkilä N, Lindbohm JV, *et al*. Hospital-treated infectious diseases and the risk of dementia: a large, multicohort, observational study with a replication cohort. *Lancet Infect Dis* 2021;21:1557–67.
8 Muzambi R, Bhaskaran K, Brayne C, *et al*. Common Bacterial Infections and Risk of Dementia or Cognitive Decline: A Systematic Review. *J Alzheimers Dis* 2020;76:1609–26.
9 Low A, Mak E, Rowe JB, *et al*. Inflammation and cerebral small vessel disease: A systematic review. *Ageing Res Rev* 2019;53:100916.
10 Faura J, Bustamante A, Miró-Mur F, *et al*. Stroke-induced immunosuppression: implications for the prevention and prediction of post-stroke infections. *J Neuroinflammation* 2021;18:127.
11 Kerkhofs D, Wong SM, Zhang E, *et al*. Blood-brain barrier leakage at baseline and cognitive decline in cerebral small vessel disease: a 2-year follow-up study. *Geroscience* 2021;43:1643–52.
12 Morton CE, Forbes HJ, Pearce N, *et al*. Association Between Common Infections and Incident Post-Stroke Dementia: A Cohort Study Using the Clinical Practice Research Datalink. *Clin Epidemiol* 2020;12:907–16.
13 Milosevich E, Demeyere N, Pendlebury ST. Infection, Inflammation, and Poststroke Cognitive Impairment. *J Am Heart Assoc* 2024;13:e9130e9130.
14 Pendlebury ST, Luengo-Fernandez R, Seeley A, *et al*. Infection, delirium, and risk of dementia in patients with and without white matter disease on previous brain imaging: a population-based study. *Lancet Healthy Longev* 2024;5:e131–40.
15 Milosevich E, Kusec A, Pendlebury ST, *et al*. Longitudinal Trajectories of Global and Domain-Specific Cognition After Stroke Using the Oxford Cognitive Screen. *Stroke* 2026;57:2150–62.
16 Milosevich E, Field TS, Pendlebury ST. ABSTRACT NUMBER: ESOC2026A1481 INFECTION AND INFLAMMATION AS RISK FACTORS FOR POST-STROKE COGNITIVE IMPAIRMENT AND DEMENTIA: A SYSTEMATIC REVIEW AND META-ANALYSIS. *Eur Stroke J* 2026;11:i66–7.
17 Boucher E, Jell A, Singh S, *et al*. Protocol for the Development and Analysis of the Oxford and Reading Cognitive Comorbidity, Frailty and Ageing Research Database-Electronic Patient Records (ORCHARD-EPR). *BMJ Open* 2024;14:e085126.
18 Bal S, Colbourne E, Gan J, *et al*. Validation of a CT-brain analysis tool for measuring global cortical atrophy in older patient cohorts. *arXiv* 2025.
19 Fazekas F, Chawluk JB, Alavi A, *et al*. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am J Roentgenol* 1987;149:351–6.
20 Wahlund LO, Barkhof F, Fazekas F, *et al*. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke* 2001;32:1318–22.
21 Pasquier F, Leys D, Weerts JG, *et al*. Inter- and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. *Eur Neurol* 1996;36:268–72.
22 Greenberg SM, Bacskai BJ, Hernandez-Guillamon M, *et al*. Cerebral amyloid angiopathy and Alzheimer disease - one peptide, two pathways. *Nat Rev Neurol* 2020;16:30–42.
23 Barber PA, Demchuk AM, Zhang J, *et al*. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. Alberta Stroke Programme Early CT Score. *Lancet* 2000;355:1670–4.
24 Goebel J, Stenzel E, Guberina N, *et al*. Automated ASPECT rating: comparison between the Frontier ASPECT Score software and the Brainomix software. *Neuroradiology* 2018;60:1267–72.
25 Pendlebury ST, Lovett NG, Thomson RJ, *et al*. Impact of a system-wide multicomponent intervention on administrative diagnostic coding for delirium and other cognitive frailty syndromes: observational prospective study. *Clin Med (Lond)* 2020;20:454–64.
26 Bone RC, Balk RA, Cerra FB, *et al*. Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis. *Chest* 1992;101:1644–55.
27 Royal College of Physicians. National early warning score (NEWS) 2: standardising the assessment of acute-illness severity in the NHS. London Updated Report of a Working Party; 2017.
28 Adams HP Jr, Bendixen BH, Kappelle LJ, *et al*. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993;24:35–41.
29 Pexman JH, Barber PA, Hill MD, *et al*. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke. *AJNR Am J Neuroradiol* 2001;22:1534–42.
30 Rockwood K, Song X, MacKnight C, *et al*. A global clinical measure of fitness and frailty in elderly people. *CMAJ* 2005;173:489–95.
31 Gilbert T, Neuburger J, Kraindler J, *et al*. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet* 2018;391:1775–82.
32 Boucher EL, Gan JM, Rothwell PM, *et al*. Prevalence and outcomes of frailty in unplanned hospital admissions: a systematic review and meta-analysis of hospital-wide and general (internal) medicine cohorts. *EClinicalMedicine* 2023;59:101947.
33 Quan H, Li B, Couris CM, *et al*. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

discharge abstracts using data from 6 countries. *Am J Epidemiol* 2011;173:676–82.
34 Boucher EL, Gan JM, Lovett NG, *et al*. Delirium prevalence, diagnostic uncertainty and outcomes in ORCHARD-EPR: validation against prospective reference cohorts. *Age Ageing* 2025;54:afaf284.
35 Boucher EL, Gan JM, Lovett NG, *et al*. Implementation of Delirium Screening at Scale in Older Patients With Emergency Hospital Admission. *JAMA Intern Med* 2025;185:884–7.
36 Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc* 1999;94:496–509.
37 Vandenbroucke JP, von Elm E, Altman DG, *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007;147:W163–94.
38 Benchimol EI, Smeeth L, Guttmann A, *et al*. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12:e1001885.
39 Collins GS, Moons KGM, Dhiman P, *et al*. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378.
40 Riley RD, Ensor J, Snell KIE, *et al*. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
41 Walsh J, Tozer DJ, Sari H, *et al*. Microglial activation and blood-brain barrier permeability in cerebral small vessel disease. *Brain* 2021;144:1361–71.

BMJ Open: first published as 10.1136/bmjopen-2026-128641 on 25 November 2026. Downloaded from http://bmjopen.bmj.com/ on November 27, 2026 at Prof K E Davies Editor, Human Molecular Genetics .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.