Public Health England

Protecting and improving the nation's health

Technical document: First incidence of stroke

Estimates for England 2007 to 2016

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. We do this through world-leading science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health and Social Care, and a distinct delivery organisation with operational autonomy to advise and support government, local authorities and the NHS in a professionally independent manner.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland

Prepared by: Tim Evans
For queries and further information relating to this document, please contact:
ncvin@phe.org.uk

Published February 2018
PHE publications
gateway number: 2017613

PHE supports the UN
Sustainable Development Goals

Contents

Glossary

CPRD:	Clinical Practice Research Datalink
GBD:	Global Burden of Disease Study
GPES:	General Practice Extraction Service
HES:	Hospital Episode Statistics
ONS:	Office for National Statistics
OXVASC:	Oxford Vascular Study
PHE:	Public Health England
QOF:	Quality and Outcomes Framework
SLSR:	South London Stroke Register
SSNAP:	Sentinel Stroke National Audit Programme
THIN:	The Health Improvement Network
TIA:	Transient ischaemic attack

Summary

Stroke incidence estimates (first ever stroke) were developed using electronic general practice records held within The Health Improvement Network database (THIN). THIN is a large general practice database covering about 6% of the UK population. The records in THIN are anonymised and longitudinal in nature. In order to compare the generalisability of the estimates in THIN, crude and age/sex standardised estimates of stroke/transient ischaemic attack (TIA) prevalence were compared to observed prevalence reported in the Quality and Outcomes Framework (QOF). This permits comparison between the estimated prevalence rate reported by GP practices and the prevalence estimates derived from THIN using the same QOF business rules.

Further validation of the method to derive national incidence included population profile comparisons between THIN, GP registered and ONS mid-year populations, distribution of prevalence estimates reported through QOF and THIN and comparison between previously reported estimates of incidence. Since the validation proved successful, THIN data were therefore used to estimate national figures.

Estimates of stroke incidence using the Read codes listed in Appendix 1, have been calculated for England between 2007 and 2016. Estimates have also been calculated by deprivation, gender, ten-year age bands and ethnicity in 2016. Stroke incidence is defined as the first ever stroke and results published here may differ from other studies that present the total number of strokes.

This paper provides technical information on the methods used to idenfity eligible patients for the analysis as well as the process for validating the estimates. All further references to incidence of stroke refer to first incidence, unless otherwise stated.

The accompanying briefing report is available to download at:
https://www.gov.uk/government/publications/first-stroke-estimates-in-england-2007-to-2016

1. Introduction

Around 85,000 stroke admissions occur each year in England, resulting in an age standardised admission rate of around 1.7 per 1,000 population in 2015/16, and there are around 32,000 stroke related deaths each year in England, with an age standardised mortality rate of 0.6 per 1,000 population in 2013/15[1]

A stroke occurs when blood supply to the brain is interrupted, usually due to blood vessels bursting or as a result of a blood clot blocking the supply of blood to the brain. As a result, the brain is starved of oxygen resulting in damage to brain tissue. Strokes can be broadly split into 2 types: ischaemic and haemorrhagic strokes. Ischaemic

strokes outnumber haemorrhagic strokes by around 8:2[2,3]. Haemorrhagic strokes are associated with increased mortality due to the actual bleeding within the brain, compared to ischaemic strokes which are caused by blockages of blood supply.

There are a number of factors that increase the risk of stroke. These include non-modifiable risks such as age, ethnicity, family history and certain genetic conditions. Modifiable risk factors associated with stroke include, high blood pressure, high cholesterol, diabetes and atrial fibrillation[4]. There are a number of lifestyle factors that contribute to some of the modifiable risk factors including smoking, excessive alcohol intake, being obese, low physical activity and eating unhealthy food. These can lead to damaged blood vessels and increase circulatory problems.

There are a variety of approaches for estimating incidence and prevalence of stroke within England. There are local and regional registries, for example the South London Stroke register which gathers data on all first ever strokes in south London, in particular Lambeth and Southwark. The Sentinel Stroke National Audit Programme (SSNAP) gathers national data on the hospital burden of stroke, including data on all strokes (incident and recurrent) among individuals who have been admitted to hospital. In addition, Hospital Episode Statistics (HES) gathers national data on stroke-related hospital admissions, including first stroke and readmission for stroke, although from HES, it is not possible to distinguish between first stroke and subsequent admissions. The Quality and Outcomes Framework (QOF) is a voluntary annual reward and incentive programme for the majority of practices in England. As part of the programme, GPs maintain registers of patients who have had a stroke or TIA and this serves as a source of prevalence data.

General practice data can be a rich source of longitudinal data on patients. Within this rich source, data relating to stroke diagnoses and treatment are collected along the patient journey. There are a number of sources of general practice data, including research databases capturing populations of England (Clinical Practice Research Datalink (CPRD)[5] and The Health Improvement Network database (THIN)); and tailor made extracts of data can be gathered via the General Practice Extraction Service (GPES)[6].

There is no population-based disease register for stroke which covers all of England; and with an absence of data on incident stroke it was necessary to produce estimated figures.

A copy of the THIN database is held by Public Health England (PHE); as a result, use and access of this source of data was preferable to obtaining other similar datasets that are available for such research. Annex 1 contains a more detailed review of data sources currently available

2. Methods

Using the available THIN dataset (THIN 1701) held by PHE, the analysis was conducted in 2 stages. First, explolarotory work was conducted to ascertain how representative data within THIN are compared to the England population and therefore how viable THIN might be as a source of incidence data for stroke. Briefly, this entailed comparing estimated QOF defined stroke and TIA prevalence with nationally published data. Following this, the distribution of GP practice level estimated prevalence within THIN was compared to the distribution of QOF published practice level estimates. Finally age profiles of THIN, GP registered populations and ONS mid-year population estimates were compared. Following positive results, the full analysis was conducted.

2.1 Description of THIN

THIN has been collecting data from general practices since 2002. It began as collaboration between the providers of the Vision software (InPractice Systems) used by some general practices and the Epidemiology and Pharmacology Information Core (EPIC) which provided primary care data for medical research[7]. At the point of first data extraction for a practice, all historic electronic data are extracted, after which, data are collected routinely every few months. Demographic data are quality assured to ensure fitness for use, while clinical data are untransformed. For some practices, electronic data collection has been in operation since 1987 meaning that around 30 years of data are available. Since data collection began, over 700 Vision practices have joined the scheme and contributed data to the THIN database[8,9].

The data consist of records from every GP consultation from participating practices. Read codes are a hierarchical coding system allowing for the easy recording of patient events, including diagnoses, symptoms, procedures and laboratory results. Data are also recorded on drug therapy prescribed to patients. The combined data are anonymised in order to protect patient privacy and access to the data is now managed by IQVIA, formerly IMS Health[10], while each research project is assessed by an independent scientific review committee[11].

2.2 Practice inclusion criteria

In order to obtain a stable cohort of contributing practices for the analysis, a practice start date was generated from the greater of: one-year post Vision installation date, acceptable mortality reporting date or 1 January 2007. The rationale for including practices one year after the installation of the Vision practice software was to allow for an implementation phase of the software. This would ensure the full capabilities of the software were being utilised. If the generated start date of the practice was greater than

1 January 2007, these practices were excluded. In a similar approach, a practice end date was generated which took the lesser of the last known collection date of data for the practice or 31 December 2016. Practices with an end date earlier than 31 December 2016 were excluded. As a result of these selection criteria, a stable cohort of 291 practices that contributed continuously too THIN between 1 January 2007 and 31 December 2016 were identified and used in this analysis.

2.3 Patient inclusion criteria

Patients were eligible for inclusion in the analysis if they had been registered at their practice for 12 months or more and were resident in England. THIN receives data from participating practices across England, Wales, Northern Ireland and Scotland. The analysis presented in the briefing document is in relation to England only.

Patients were only included if they had a complete registration date recorded in THIN and met the following criteria:

1. Patient flag was either ‘A’ or ‘C’ indicating that the patient record had passed the internal THIN validation process and was an ‘acceptable record’ or was an ‘acceptable record and transferred out of data due to patient being dead’.
2. Registration flag was either ‘01’, ‘02’, ‘05’ or ‘99’ indicating that a patient has ‘Applied’, or ‘Permanent’ residence or patient has ‘transferred out’ of the practice or patient has died’.
3. Patient start date was earlier than 31 December 2016 and patient start date was earlier than the patient end date.

In order to identify stroke patients in the incidence analysis, patient records were searched for the earliest mention of an occurrence of stroke according to a defined lookup file containing Read codes. Where this occurred outside the study period, they were classed as prevalent cases. Where patients had 2 or more stroke records within the study period, the earliest record was used as the first incidence, with subsequent records classed as prevalent cases. Where the only record in the patient record suggested a history of stroke, these too were classed as prevalent only cases. Once classified, all records identified as prevalent cases were excluded from the incidence analysis. Patients with a Read code indicating stroke, but with missing diagnosis dates were also excluded Without a definitive date of diagnosis it is not possible to determine when this event occurred. This may result in a small under-reporting of incidence rates.

The lookup file of Read codes for including diagnoses of stroke were reviewed by medically qualified colleagues and updated where necessary. The analysis was conducted on the full population and therefore no patient was excluded on the basis of

age. Any results published should be indicative and generalisable to the population of England.

The analysis reported in the accompanying briefing document to this technical report has been presented by year, gender, age, deprivation and ethnicity. Exact age is not provided within THIN; therefore this was estimated using year of birth provided in THIN. Diagnosis year was subtracted from year of birth to give an approximate age at onset of first stroke. A patient's deprivation quintile for each patient is provided in THIN by linking patient postcode at point of data collection. The Townsend index is used in THIN to determine deprivation with increasing deprivation as quintile increases. As patients may move house during the study period, and therefore have multiple deprivation records, the deprivation quintile assigned at the time of stroke onset was used in the analysis. Ethnicity and patient gender are recorded directly at GP practices; in order to avoid the potential of patients having multiple ethnicities, the last recorded value was retained for the analysis.

Data were extracted from the 1701 version of the THIN database and the study was approved by the Scientific Review Committee overseeing research using THIN data.

2.4 Validation of approach

To assess the validity of the stroke incidence estimates from THIN, a comparison was made between the prevalence of stroke and TIA reported in QOF and the prevalence of stroke and TIA within the THIN population. The same inclusion criteria were used for including patients as well as mirroring the QOF Read code business rules. Comparisons were also made of the distribution of prevalence by the practices within THIN and QOF to determine the similarly between data sources. Finally, age profiles of the THIN population, GP registered populations and mid-year population estimates for residents in England were compared. The comparisons of prevalence distribution and population profiles are shown in Figures 1 – 3 below. The comparisons show that it is possible to derive nationally similar prevalence estimates from the THIN data due to similarities in prevalence and age profiles from nationally available data sources.

In order to compare age structure of the population in THIN, general practice populations were obtained from NHS Digital[12], as were practice level estimates of stroke/TIA prevalence as reported through QOF[13]. Single year of age populations split by gender were also obtained from the Office for National Statistics[14]. In order to assess prevalence of stroke/TIA in THIN compared to QOF, Read codes for comparing prevalence (including 'incident', 'history of' and 'prevalent' stroke) were identified from QOF business rules version 34[15] and applied to relevant records in THIN in order to compare prevalence of stroke/TIA in THIN and QOF.

Since the stroke and TIA prevalence estimates and age profiles of THIN data were comparable to externally validated whole population data sources, it is assumed that any stroke incidence estimates will be similarly representative.

3. Statistical analysis

Incidence rates were calculated by age, sex and calendar year. Crude rates were simply a sum of the identified cases divided by the sum of the total population at risk during the period of interest. Age-standardised incidence rates were calculated by applying the 2013 European standard population distribution to our study population. A comparison was made between standardising the rates to the 2013 European standard population and the national population for England was undertaken which showed little variation between approaches. In order to maximise international comparisons, only the 2013 European age-standardised rates are presented.

To estimate the total number of strokes in England, age and gender specific incidence rates from THIN were applied to the English national population obtained from ONS. The number of cases in each age and gender group was summed to provide an overall number.

All data manipulation and analysis was undertaken using Stata (StataCorp. 2015. Stata Statistical Software: Release 14.1. College Station, TX: StataCorp LP).

Figure 1: Stroke and TIA prevalence observed from QOF data and THIN

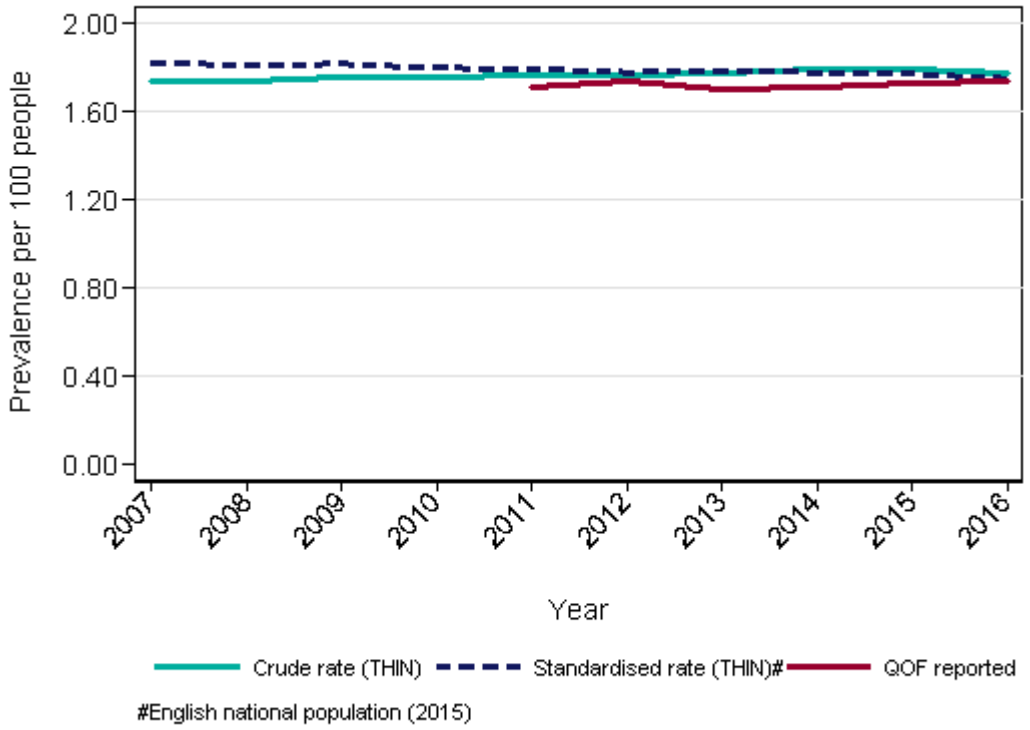


Figure 2: Stroke and TIA prevalence distribution by QOF general practice and THIN general practice, England

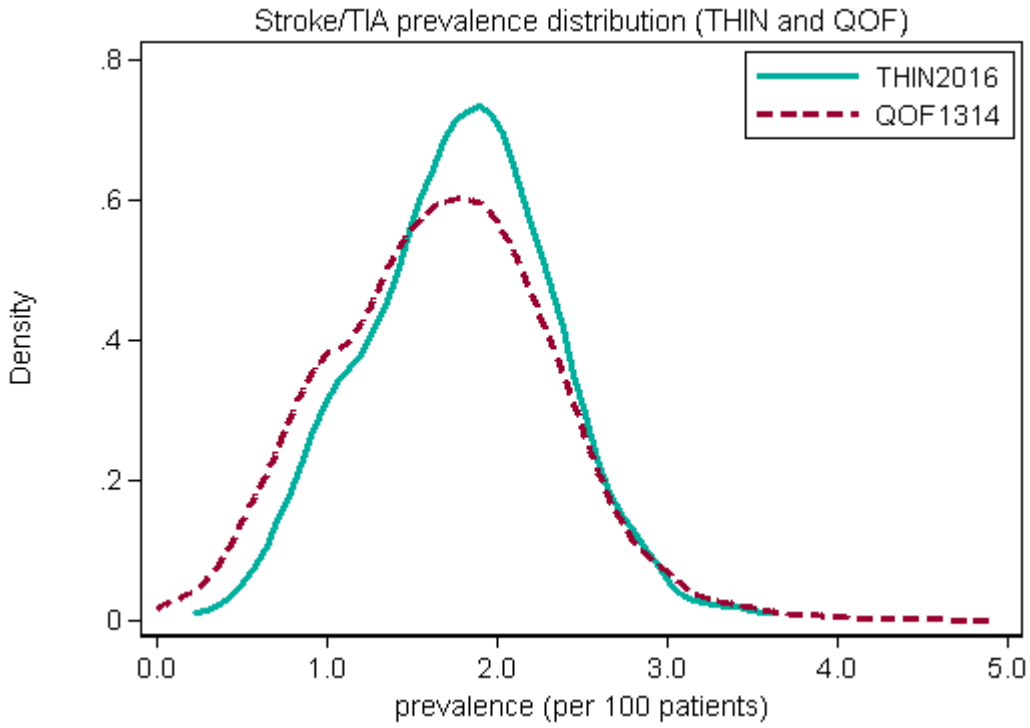
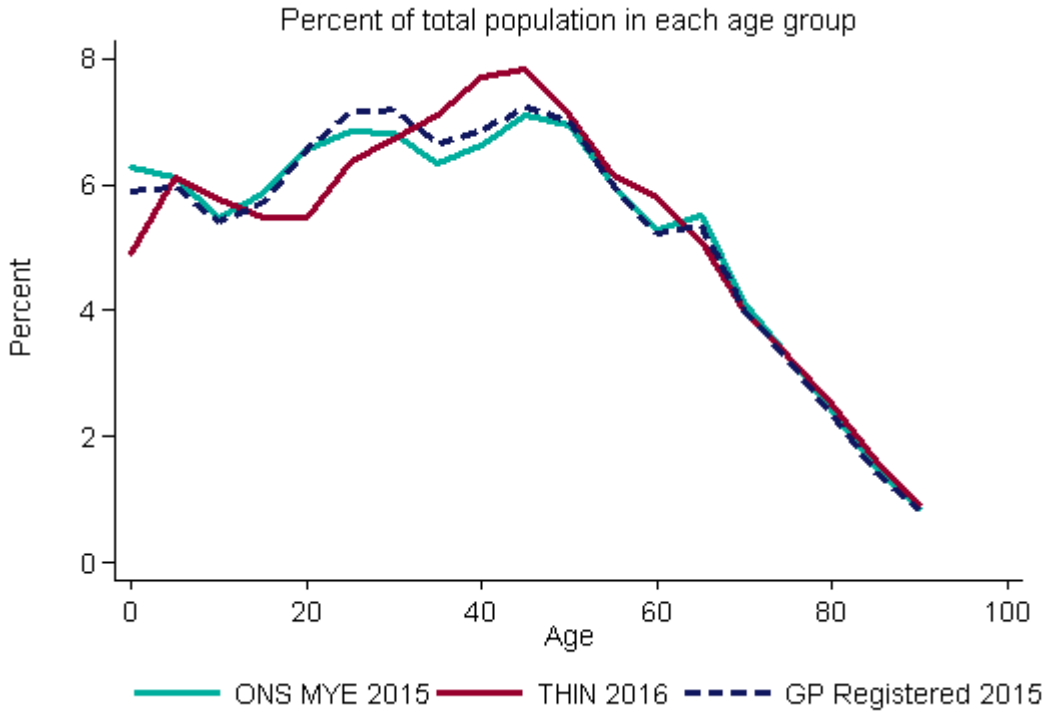


Figure 3: Age distribution of THIN, general practice populations and mid-year resident populations



4. Strengths and Limitations

There are a number of strengths around the use of THIN in producing population level estimates:

1. Data collected in THIN includes both primary and secondary care data, meaning out-of-hospital health events are more likely to be captured.
2. THIN contains data from 6% of the population of England; this is a large database, and it has been shown to be representative of the general population.
3. Data allows for the selection of all patients with a particular condition to be studied, subject to that condition being recorded in the data.
4. Data are continuously updated, allowing for robust time-series analysis of events in both the cohort and individual patients. The data can be used for longitudinal research allowing for long-term follow-up of patients.
5. THIN has been used extensively in other epidemiological studies, for example in illuratrating the prescribing of statins for the primary prevention of CVD, frequency of antibiotic prescribing as well as under use of prevention drugs in the prevention of strokes and TIA's.

There are also a number of limitations to bear in mind in interpreting the results.

1. The method assumes the epidemiology of stroke within the limited number of practices participating in THIN accurately reflects both behavioural and fixed risk factors for stroke in the wider population in England.
2. Only the first ever incidence of stroke can be accurately estimated since it is not possible to be completely certain that subsequent coding of strokes data within the GP record do relate to ongoing care of the initial stroke, or are in fact new incident cases.
3. Extrapolated estimates from THIN assume that case ascertainment is good and GP practices will receive all health information for patients from all relevant medical sources.
4. It is not possible to assess the missing data from THIN and therefore it is assumed any missing data are missing at random. In mitigation of potential missing data, GPs have for a long time been incentivised to regularly collect healthcare data.
5. There is an under-representation of practices from deprived areas within the THIN database.
6. The incidence rates published are estimates, and without verified national baseline data, it is difficult to understand how close to the truth these estimates are. However, when compared against other published incidence estimates they are comparable.

Email the National Cardiovascular Network (NCVIN) for further details: ncvin@phe.gov.uk

Appendix 1: Read codes for stroke incidence

Table A1: Read codes for stroke incidence

Read code	Description
G61..00	intracerebral haemorrhage
G61..11	cva - cerebrovascular accid due to intracerebral haemorrhage
G61..12	stroke due to intracerebral haemorrhage
G610.00	cortical haemorrhage
G611.00	internal capsule haemorrhage
G612.00	basal nucleus haemorrhage
G613.00	cerebellar haemorrhage
G614.00	pontine haemorrhage
G615.00	bulbar haemorrhage
G616.00	external capsule haemorrhage
G617.00	intracerebral haemorrhage, intraventricular
G618.00	intracerebral haemorrhage, multiple localized
G619.00	lobar cerebral haemorrhage
G61X.00	intracerebral haemorrhage in hemisphere, unspecified
G61X000	left sided intracerebral haemorrhage, unspecified
G61X100	right sided intracerebral haemorrhage, unspecified
G61z.00	intracerebral haemorrhage NOS
Gyu6200	[x]other intracerebral haemorrhage
Gyu6F00	[x]intracerebral haemorrhage in hemisphere, unspecified
G66..11	cva unspecified
G66..12	stroke unspecified
G66..13	cva - Cerebrovascular accident unspecified
G667.00	left sided cva
G668.00	right sided cva
G60..00	subarachnoid haemorrhage
G600.00	ruptured berry aneurysm
G601.00	subarachnoid haemorrhage from carotid siphon and bifurcation
G602.00	subarachnoid haemorrhage from middle cerebral artery
G603.00	subarachnoid haemorrhage from anterior communicating artery
G604.00	subarachnoid haemorrhage from posterior communicating artery
G605.00	subarachnoid haemorrhage from basilar artery
G606.00	subarachnoid haemorrhage from vertebral artery
G60X.00	subarachnoid haemorrh from intracranial artery, unspecif
G60z.00	subarachnoid haemorrhage nos
Gyu6000	[x]subarachnoid haemorrhage from other intracranial arteries

Table A1: Read codes for stroke incidence (continued)

Read code	Description
Gyu6100	[x]other subarachnoid haemorrhage
Gyu6E00	[x]subarachnoid haemorrh from intracranial artery, unspecif
G63y000	cerebral infarct due to thrombosis of precerebral arteries
G63y100	cerebral infarction due to embolism of precerebral arteries
G64..00	cerebral arterial occlusion
G64..11	cva - cerebral artery occlusion
G64..12	infarction - cerebral
G64..13	stroke due to cerebral arterial occlusion
G640.00	cerebral thrombosis
G640000	cerebral infarction due to thrombosis of cerebral arteries
G641.00	cerebral embolism
G641.11	cerebral embolus
G641000	cerebral infarction due to embolism of cerebral arteries
G64z.00	cerebral infarction NOS
G64z.11	brainstem infarction NOS
G64z.12	cerebellar infarction
G64z000	brainstem infarction
G64z100	wallenberg syndrome
G64z111	lateral medullary syndrome
G64z200	left sided cerebral infarction
G64z300	right sided cerebral infarction
G64z400	infarction of basal ganglia
G66..00	stroke and cerebrovascular accident unspecified
G660.00	middle cerebral artery syndrome
G661.00	anterior cerebral artery syndrome
G662.00	posterior cerebral artery syndrome
G663.00	brain stem stroke syndrome
G664.00	cerebellar stroke syndrome
G665.00	pure motor lacunar syndrome
G666.00	pure sensory lacunar syndrome
G676000	cereb infarct due cerebral venous thrombosis, nonpyogenic
G6W..00	cereb infarct due unsp occlus/stenos precerebr arteries
G6X..00	cerebrl infarctn due/unspcf occlusn or sten/cerebrl artrs
Gyu6300	[x]cerebrl infarctn due/unspcf occlusn or sten/cerebrl artrs
Gyu6400	[x]other cerebral infarction
Gyu6500	[x]occlusion and stenosis of other precerebral arteries
Gyu6600	[x]occlusion and stenosis of other cerebral arteries
Gyu6G00	[x]cereb infarct due unsp occlus/stenos precerebr arteries

Appendix 2: Review of literature on previous estimates of stroke incidence

The most recent estimate of the number of strokes suggests that annually in the UK there are over 100,000[16]. This would equate to an incidence rate of around 1.5 per 1,000 people. If this rate were applied to England, then it would be anticipated that around 81,000 strokes occur each year. These figures are based upon data obtained from the Sentinel Stroke National Audit Programme[17] which tracks patients who have had a stroke and been admitted to hospital. These figures may be subject to uncertainty since case-ascertainment of the audit may not be complete[18], and does not include any patient that has had a stroke and not been admitted to hospital. The total figures may also differ to other published studies because these figures represent the total number of strokes, rather than the number of first strokes admitted to hospital.

The OXVASC (Oxford Vascular) study, established in 2002[19] is an ongoing population-based study that covers around 93,000 residents in 9 general practices in Oxford, England[20] and was established primarily to identify the causes and commonality of vascular disease including strokes, TIAs and other circulatory problems. Its strengths are that it includes all registered residents of the general practices collaborating in the study. Around 97% of the Oxfordshire population is registered with a GP and therefore this study can be assumed to cover the whole population of those practices[21]. The broad ethnic and socio-economic make-up of OXVASC is similar to England, although there will be significant differences on a region by region basis that are not represented in the OXVASC study. The most recent estimates from this study from 2004 showed that the age standardised incidence of stroke in 2002/04 was 1.6 strokes per 1,000 population. Females experienced higher incidence rates compared to males: 1.7 strokes per 1,000 males per year compared to 1.5 strokes per 1,000 females.

The South London Stroke Register (SLSR) is also another population-based registry, focusing on the prospective collection of all first ever strokes in the populations of Lambeth and Southwark in South London[22]. This is a varied population with a diverse ethnic population where more than 35% of the study source population are classed as non-white[23]. The population covered in the SLSR is significantly different to the England population where over 86% of the population identify themselves as white[24]. When there are differences in risk factors and incidence of stroke as determined by ethnicity, it may not be feasible to extrapolate national estimates from such a diverse population. The most recent estimates for stroke incidence reported by the SLSR for the period 2003/04 showed the same differential incidence pattern between males and females, but with a slightly reduced incidence of 1.3 strokes per 1,000 males and 0.9 strokes per 1,000 females. The difference in gender specific incidence rates between SLSR and

OXVASC could be a result of demographic differences in the populations, or the definition of stroke used in the 2 studies.

The Scottish Borders Stroke Study was set up as a community-based study of stroke and TIA carried out between 1998 and 2000[25]. The study area was noted for its rural base and a slightly older population compared to the general population. First ever incidence of stroke was reported higher than other studies, with observed crude rates being 2.8 strokes per 1,000 population. This is twice the rate reported in most other UK studies, although the higher incidence was attributed to the higher average age of the population. This age factor was further emphasised when the data were adjusted to reflect the demographic spread of the population, with adjusted rates being comparable to other incidence studies.

Perhaps the most recent estimates of stroke incidence in the UK to be published were from the Global Burden of Disease Study (2010). Feigin *et al*[26] undertook a meta-analysis of existing published studies before applying the GBD analytical technique (DisMod-MR). The GBD technique works by applying disease specific and pre-determined associations between incidence, mortality and prevalence of a given disease. The associations are then used to calculate age, region, income level and country specific disease burdens including incidence. In this study stroke incidence for the UK was estimated to be 1.2 strokes per 1,000 population. These are in line with other UK based studies, and the good agreement is perhaps unsurprising for the UK as the main sources for calculating incidence specifically for the UK would have been OXVASC, the Scottish Borders Stroke Study and the South London Stroke Register. The GBD is updated annually and includes sub-national estimates of disease for England and smaller areas. The most recent estimates of incidence for England are available from their online database and indicate that more than 110,000 strokes occurred in 2016[27].

Use of primary care data is not a new phenomenon in epidemiological research studies. Such databases have been used to assess the incidence and prevalence of a variety of conditions, including Huntingdon's disease[28], Coronary Heart Disease[29], diabetic retinopathy[30], skin cancer[31], chronic kidney disease[32] and stroke[33]. The study by Lee *et al*[34] using GPRD (now known as CPRD) showed that over a ten year period, incidence of stroke had decreased by 30% between 1999 and 2008. Stroke prevalence increased during the same period from 6.4% to 7.2%. As a result of increased awareness of risk factors and improved clinical management of patients at risk of stroke, the numbers of strokes and incidence rates have declined. This might also indicate outcomes for patients who have had a stroke have improved with more patients surviving their stroke rather than dying as a result of their stroke. What is not clear from the data is whether improved outcomes are restricted to whether a patient merely survives a stroke, or whether stroke severity is also on a downward trend.

While stroke incidence continues to fall, there are still opportunities to improve this further, by identifying and addressing common risk factors to stroke. There are a number of risk factors than can increase the likelihood of a person developing atrial fibrillation (AF), these include, age[35,] ethnicity[36,37] diabetes[38,35], hypertension[39] , obesity[40] and congestive heart failure [35]. Many of the risk factors for AF are also common to stroke.

Recent estimates of atrial fibrillation prevalence suggest that around 30% of patients remain undiagnosed and unrecorded on AF registers[41]. Not only is atrial fibrillation a major risk factor in strokes, but patients who have a stroke with unmanaged AF often experience more severe strokes leading to greater disability and death. Where patients receive appropriate anticoagulation therapy, risk of stroke can be reduced by over 60%[42]. In the study by Lee *et al*[43], it was found that only 25% of patients with AF were prescribed anticoagulants before their stroke. There is therefore, considerable opportunity to reduce the incidence of stroke by firstly improving the detection of AF, but also in the subsequent treatment and management of AF.

Differences in reported crude incidence can be a result of differences in the definitions of stroke, the age and gender distributions of populations and the ethnic diversity of a given population. Given the similarity of age adjusted incidence of stroke between a number of studies, it is likely that the main driver of increased stroke burden is primarily driven by the age of the population rather than any differences in stroke risk factors.

Appendix 2.1: Approaches to calculating disease incidence

Not all of the approaches listed above are appropriate for calculating national stroke incidence. There are regional registries, for example the South London Stroke register, for which it is possible to extrapolate estimates to larger geographies; however this may risk population bias. The potential bias from the study cohort could result in either an over or underestimation of the incidence of stroke. National audit programmes are helpful at identifying hospital burden of disease; however, there are often restrictions on those patients eligible to be recorded in the audit. In the case of SSNAP, it only records patients who are actually admitted to hospital and therefore excludes those who have either died as a result of their stroke and not been admitted, or those strokes which are minor in severity and for which no secondary care is sought. Data presented from SSNAP is primarily focussed upon the total number of strokes being admitted for inpatient care, therefore any comparisons between first incidence of stroke and SSNAP data are likely to result in a large discrepancy in the number of strokes presented.

Estimates of incidence could be derived from Hospital Episode Statistics (HES), although it is not possible to determine whether an admission for stroke is related to a new episode or is a readmission due to complications from an earlier stroke. Therefore hospital admissions may overestimate the true incidence of stroke. To overcome this, it might be possible to identify the first admission for stroke and use this as a proxy for first incidence. As is the case with hospital based audits, HES admissions omit any stroke patient that fails to present to hospital.

Health surveys can capture the burden of certain diseases, although this is often presented as an overall burden in terms of prevalence, such examples include high blood pressure, adult smoking and obesity[44]. It can be the case with surveys that definitions of disease may be modified between surveys, methods of data collection may be varied and diseases can be dropped from surveys, making time-series analysis difficult. They are also limited in the number of respondents to the survey.

QOF data on diseases, like health surveys is usually presented in terms of prevalence and therefore cannot give a true reflection of yearly changes in disease incidence. National mortality statistics offer another source of disease data. This source of data could be used as a proxy for incidence where survival from a particular disease is very short and there is a lack of incidence data. However, while survival time for stroke patients does vary, many patients often experience long-term survival. Using mortality as a proxy for stroke incidence would not be an accurate measure of incidence for this particular condition and gives no indication of the effect of stroke on stroke survivors.

General practice data can be a rich source of longitudinal data on patients, especially in the era of electronic capture. There are a number of sources of general practice data, including the Clinical Practice Research Datalink (CPRD)[45], THIN and tailor made extracts of data from the General Practice Extraction Service (GPES)[46]. While these sources often only capture a small proportion of the general population, in the case of THIN and CPRD, this still accounts for around 6% of the total population. They are usually representative of the general population and have great power in terms of enabling a greater understanding of public health.

---

[1] Office for National Statistics (ONS) Annual mortality statistics. 2017 Available from <www.ons.gov.uk>.
[2] Butland, K, Atkinson, R, Crichton, S. *et al*. 2017, 'Air pollution and the incidence and haemorrhagic stroke in the South London Stroke Register: a case-cross over analysis'. *Journal of Epidemiology & Community Health*. vol. 71, no. 7, p707-712.
[3] Intercollegiate Stroke Working Party, 2016, 'National clinical guideline for stroke, 5th edition. London: Royal College of Physicians', accessed 24 October 2017, <https://www.strokeaudit.org/SupportFiles/Documents/Guidelines/2016-National-Clinical-Guideline-for-Stroke-5t-(1).aspx>.
[4] O'Donnell, M, Xavier, D, Lisheng, L. *et al*. 2010, 'Risk factors for ischaemic and intracereberal haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study'. *The Lancet*. vol. 376, p112-123.
[5] Clinical Practice Research Datalink (CPRD) access 26 October 2017, < https://www.cprd.com/home/>.
[6] NHS Digital. 'General Practice Extraction Service' accessed 26 October 2017, < https://digital.nhs.uk/General-Practice-Extraction-Service>.
[7] Tagger, J, Coleman, T, Lewis, S. *et al*. 2012, 'The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database'. *BMC Public Health*. vol.12, no. 329.
[8] University College London, 'The Health Improvement Network (THIN) research team – THIN Database', accessed 7 November 2017, < https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database>.
[9] Denburg, M, Haynes, K, Shults, J, *et al*. 2011, Validation of The Health Improvement Network (THIN) database for epidemiological studies of chronic kidney disease'. *Pharmacoepidemiology and Drug Safety.* vol. 20, no. 11, p1138 – 1149.
[10] IQVIA, accessed 7 November 2017, < https://www.iqvia.com/about-us>.
[11] QuintilesIMS. 2017, 'THIN Data Guide for Researchers', version 1701.
[12] NHS Digital, 'Number of Patients Registered at a GP practice – Oct 2016', accessed 10 September 2017, <http://www.content.digital.nhs.uk/catalogue/PUB22008/gp-reg-patients-prac-quin-age.csv>.
[13] NHS Digital, 'Quality and Outcomes Framework (QOF) – 2015-16', accessed 17 October 2017, < http://digital.nhs.uk/catalogue/PUB22266>.
[14] Office for National Statistics, 'Middle Super Output Area Mid-Year Population Estimates', accessed 29 September 2017, <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/middlesuperoutputareamidyearpopulationestimates>.
[15] NHS Digital, 'New GMS Contract QOF Implementation Dataset and Business Rules - Stroke and Transient Ischaemic Attack (STIA) Indicator Set', accessed 24 October 2017, <http://content.digital.nhs.uk/media/21023/Stroke-and-Transient-Ischaemic-Attack-rulesetv340/pdf/Stroke_and_Transient_Ischaemic_Attack_ruleset_v34.0.pdf>.
[16] Stroke Association, 'State of the Nation: Stroke statistics', accessed 19 October 2017, <https://www.stroke.org.uk/sites/default/files/state_of_the_nation_2017_final_1.pdf>.
[17] Royal College of Physicians Sentinel Stroke National Audit Programme (SSNAP). 'Is stroke care improving?' Second SSNAP Annual Report prepared on behalf of the Intercollegiate Stroke Working Party, November 2015.
[18] Bray, B, Cloud, G, James, M. *et al*. 2016, 'Weekly variation in health-care quality by day and time of admission: a nationwide, registry-based, prospective cohort study of acute stroke care', *The Lancet*. vol. 388, no. 10040, p170-177.
[19] Oxford Vascular Study, accessed 19 October 2017, <https://www.ndcn.ox.ac.uk/research/oxvasc>.

[20] Webb, A, Wilson, M, Lovett, N. *et al*. 2014, 'Response of Day-to-Day Home Blood Pressure Variability by Antihypertensive Drug Class After Transient Ischemic Attack or Nondisabling Stroke'. *Stroke.* vol. 45, p2967-2973.
[21] Rothewell, P, Coull, A, Silver, L. *et al*. 2005, 'Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study)'. *The Lancet*. vol. 366, no. 9499, p1773-1783.
[22] Addo, J, Bhalla, A, Crichton, S. *et al*. 2011, 'Provision of acute stroke care and associated factors in a multi-ethnic population: prospective study with the South London Stroke Register'. *BMJ.* vol. 342, no. 744.
[23] Hajat, C, Heuschmann, P, Coshall, C. *et al*. 2011, 'Incidence of aetiological subtypes of stroke in a multi-ethnic population based study: The South London Stroke Register'. *Journal of Neurology, Neurosurgery & Psychiatry.* vol. 82, p527-533.
[24] Office for National Statistics (ONS), 'Ethnicity and National Identity in England and Wales: 2011' accessed 12 October 2017, <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11>.
[25] Syme, P, Byrne, A, Chen, R. *et al.* 2005, 'Community-based stroke incidence in a Scottish population: The Scottish Borders Stroke Study'. *Stroke*. vol. 36, no. 9. p1837-1843.
[26] Feigin, V, Forouzanfar, M, Krishnamurthi, R. *et al.* 2014, 'Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010'. *The Lancet.* vol. 383. p245-255.
[27] Institute for Health Metrics and Evaluation (IMHE), University of Washington, 'Global Health Data Exchange – GBD Results Tool', accessed 26 October 2017, <http://ghdx.healthdata.org/gbd-results-tool>.
[28] Wexler, N, Collett, L, Wexler, A. *et al* . 2016, 'Incidence of adult Huntingdon's disease in the UK: A UK based primary care study and a systematic review'. *BMJ Open*. vol. 6, no. 2.
[29] Bhattarai, N, Charlton, J, Rudisill, C. *et al*. 2012, 'Coding, recording and incidence of different forms of Coronary Heart Disease in Primary Care'. *PLoS ONE*. vol. 7, no. 1.
[30] Mathur, R, Bhaskaran, K, Edwards, E. *et al.* 2017, 'Population trends in the 10-year incidence and prevalence of diabetic retinopathy in the UK: a cohort study in the Clinical Practice Research Datalink 2004 – 2014'. *BMJ Open*. vol. 7, no. 2.
[31] Bath-Hextall, F, Leonardi-Bee, J, Smith, C. *et al.* 2007, 'Trends in incidence of skin basal cell carcinoma. Additional evidence from a UK primary care database study'. *International Journal of Cancer*. vol. 121, p2105-2108.
[32] Iwagami, M, Tomlinson, L, Mansfield, K. *et al*. 2017, 'Validity of estimated prevalence of decreased kidney function and renal replacement therapy from primary care electronic health records compared with national survey and registry data in the United Kingdom'. *Nephrology Dialysis Transplantation*. vol. 32, p.ii142-ii150.
[33] Lee, S, Shafe, A & Cowie, M. 2011, 'UK stroke and incidence, mortality and cardiovascular risk management 1998-2008. Time trend analysis from General Practice Research Database'. *BMJ Open*. vol. 1, no. 2.
[34] Lee, S, Shafe, A & Cowie, M. 2011, 'UK stroke and incidence, mortality and cardiovascular risk management 1998-2008. Time trend analysis from General Practice Research Database'. *BMJ Open*. vol. 1, no. 2.
[35] Benjamin, E, Levy, D, Vaziri, S. *et al.* 1994, 'Independent Risk Factors for Atrial Fibrillation in a Population-Based Cohort: The Framingham Heart Study'. *JAMA.* vol. 271, no. 11, p840-844.
[36] Davis, R, Hobbs, F, Kenkre, J. *et* al. 2012, 'Prevalence of atrial fibrillation in the general population and in high-risk groups: the ECHOES study'. *Europace.* vol. 14, no. 11, p1553-1559.
[37] Shen, A, Contreras, R, Sobnosky, S. *et al*. 2010, 'Racial/ethnic differences in the prevalence of atrial fibrillation among older adults - a cross sectional study'. *Journal of the National Medical Association.* vol. 102, no. 10, p906-913.
[38] Dublin, S, Glazer, N, Smith, N. *et al*. 2010, 'Diabetes Mellitus, Glycemic Control, and Risk of Atrial Fibrillation'. *Journal of General Internal Medicine*. vol. 25, no. 8, p.853-858.
[39] Schoonderwoerd, B, Smit, M, Pel, L. *et al*. 2008, 'New risk factors for atrial fibrillation: causes of 'not-so-lone atrial fibrillation''. *Europace* . vol. 10, no. 6, p.668-673.
[40] Watanabe, H, Tanabe, N, Watanabe, T. *et al*. 2008, 'Metabolic syndrome and risk of development of atrial fibrillation: the Niigata preventive medicine study'. *Circulation*. vol. 117, no. 10, p1255-1260.
[41] Public Health England. 2017, 'Atrial fibrillation prevalence estimates in England: Application of recent population estimates of AF in Sweden', accessed 10 October 2017, <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/644869/atrial_fibrillation_AF_briefing.pdf>.
[42] Katritsis, D, Gersh, B & Camm, A. 2015, 'Anticoagulation in Atrial Fibrillation – Current Concepts'. *Arrhythmia & Electrophysiology Review*. vol. 4, no. 2, p100-107.

[43] Lee, S, Shafe, A & Cowie, M. 2011, 'UK stroke and incidence, mortality and cardiovascular risk management 1998-2008. Time trend analysis from General Practice Research Database'. *BMJ Open*. vol. 1, no. 2.
[44] NHS Digital. 2015. 'Health Survey for England 2015: Health, social care and lifestyles – summary of key findings, accessed 20 October 2017, < http://digital.nhs.uk/catalogue/PUB22610>.
[45] Clinical Practice Research Datalink (CPRD) access 26 October 2017, < https://www.cprd.com/home/>.
[46] NHS Digital. 'General Practice Extraction Service' accessed 26 October 2017, < https://digital.nhs.uk/General-Practice-Extraction-Service>.