Does Ethnicity Alters the Risk of Stroke in Patients with Modifiable Cardiometabolic Risk Factors?

Marc Delord[1,*], Mark Ashworth[1], and Abdel Douiri[1]

[1]School of Life Course & Population Sciences, Department of Population Health Sciences, King's College London, London, United Kingdom
[*]Correspondence: marc.delord@kcl.ac.uk, Tel: +44 20 7848 8710

Abstract

Background: Ethnic inequality in stroke risk was partially explained by history of hypertension, diabetes, and socioeconomic status. We aimed to estimate the impact of ethnicity on stroke risk in patients with hypertension, diabetes or hypercholesterolaemia.

Methods: Multistate models were constructed using electronic health records including date at onset of stroke, stroke risk factors, and the Index of Multiple Deprivation in adult patients of Asian, Black, or White ethnicity, registered in 41 general practices in south London between 2005 and 2021. Transitions to hypertension, diabetes, or hypercholesterolaemia (transition 1) and from these risk factors to stroke (transitions 2) were considered.

Results: Of the 849,968 registered patients, 651,888 (76·7%) were of Asian (8·6%), Black (18·2%), or White (73·2%) ethnicity and 54·1% were female. Patients of Black ethnicity had higher incidence of stroke (2·1%) and cardiometabolic risk factors, whereas patients of Asian and White ethnicity had intermediate and lower levels of stroke (1·2% and 1·0% respectively) and cardiometabolic risk factors. Asian and Black ethnicities were associated with higher risks in transition 1 compared to White ethnicity. Ethnicity did not alter transition 2, except for hypercholesterolaemia, Black ethnicity being associated with higher risk of stroke compared to Asian and White ethnicity.

Conclusion: Our results confirm that Black ethnicity does not confer excess stroke risk in patients with hypertension or diabetes. This observation was also made for Asian ethnicity. Black ethnicity was associated with increased stroke risk only in patients with Hypercholesterolaemia.

Funding: Guys and St Thomas grant (EIC180702).

Keywords:

Ethnic disparity; Stroke; Cardiometabolic risk factors; hypercholesterolaemia; Electronic Health Records; Primary Health Care; Life Course Epidemiology

Introduction

The excess in stroke incidence in populations of Black versus White ethnicity is well documented [1, 2, 3, 4, 5]. This excess has been associated with higher prevalence of modifiable cardiometabolic stroke risk factors in patients of Black ethnicity, especially hypertension and diabetes [1]. The extent to which these risk factors account for this excess risk of stroke in patients of Black ethnicity remains however controversial, even in studies which found that the Black versus White ethnic disparity could not be fully explained by hypertension and diabetes alone. Yet, the addition of socioeconomic confounders such as levels of income or education generally provide further insight into observed inequalities [3, 6]. Discrepancies between studies suggest that alternative factors such as childhood socioeconomic status [3, 7], genetic / epigenetic factors [8, 9, 10], or even the potency of traditional risk factors may still explain part of this difference.

Data on the incidence of stroke and prevalence of associated risk factors among patients of Asian ethnicity are less abundant. Available studies report higher incidence of stroke and risk factor profiles similar to those observed in patients of Black ethnicity, i.e. characterised by high prevalence of hypertension, cardiac diseases and notably diabetes. However, the sociodemographic profiles of these patients resemble those of White ethnicity with little social gradient reported between ethnic groups [11, 12].

The stroke potency of modifiable cardiometabolic risk factors (hypertension, diabetes, and hypercholesterolaemia) according to different ethnic background remains poorly understood. This question can be effectively addressed using population based longitudinal electronic health records in the framework of multistate models [13]. Using this class of model, we aim to evaluate the impact of ethnicity on the hazard of stroke modifiable cardiometabolic risk factors and on the hazard of stroke mediated by either hypertension, diabetes, or hypercholesterolaemia in a young, multiethnic, urban population.

Method

Study participants

The study was conducted in the Lambeth borough in South London, characterised by a multiethnic, youthful, and deprived population. We utilised primary care electronic health records from adult patients aged over 18 and registered in one of the 41 Lambeth general practices between April 2005 and April 2021. 3·2% of patients from the registry dropped out from the anonymized data sharing scheme and therefore were excluded from the study.

Primary care data

Electronic health records included date at diagnoses of 18 long term conditions coded in accordance with the Quality and Outcomes Framework (QOF), based on QOF38 definitions.

Long term conditions included stroke, hypertension, diabetes, hypercholesterolaemia (total cholesterol > 5·0 mmol/L), atrial fibrillation, heart failure, and myocardial infraction. Other information included gender (self ascribed), ethnicity (Asian ethnicity, Black ethnicity, mixed ethnicity, other/unknown ethnicity, and White ethnicity), 'ever smoking' status, and the Index of Multiple Deprivation (IMD) 2019 (from 1-most deprived to 5-least deprived, based on seven domains of deprivation including income, employment, education, health, crime, housing, and quality of living environment). The study were conducted on patients of Asian, Black and White ethnicity.

Stroke risk factors

For consistency with previous studies [1, 3, 6] and dependent upon data availability, we have restricted our analysis to risk factors related to the traditional risk factors, demographic variables as well as a measure of the socioeconomic status. These include gender, hypertension, diabetes, hypercholesterolaemia, 'ever smoking' status, cardiovascular disease history (prior myocardial infarction, coronary heart disease, and heart failure), atrial fibrillation, and the IMD 2019. Of note, age, which is implicitly accounted for through the dating of records for long term conditions, was not added as a separate variable.

Multistate models and statistical analyses

Multistate models used in this study, are extensions of survival models were multiple time-to-even variables can be analysed in association with covariates. This allows a more focused insight into the biological and clinical processes leading to diseases by evaluating the impact of prognostic factors in different phases of diseases processes. Classical examples include the disease/recovery process or the progressive illness/death model where intermediate events such as the occurrence of a long term condition can influence the probability of subsequent events.

The progressive illness/death multistate model can be adapted to the setting of stroke where a cardiometabolic risk factor represent an intermediate state that patients may visit before eventually experiencing stoke. This progressive risk factor/disease model can be enriched by adding death as a semi competing event for stroke and other relevant transitions such as shown in figure 1.

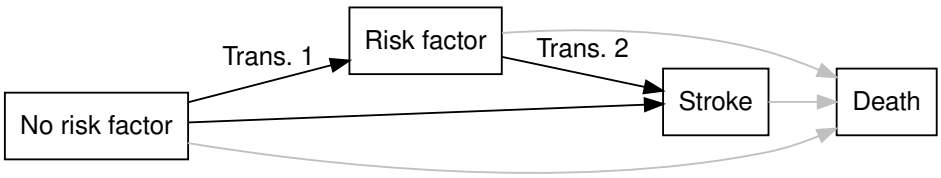

Figure 1: *Progressive risk factor/disease model for stroke. The progressive risk factor/disease model is represented by black arrows. Grey arrows represent other estimable cause specific hazards*

In this study, we analysed sequentially hypertension, diabetes, and hypercholesterolaemia as risk factors for stroke according to the diagram in figure 1. Transition-specific hazard ratios were computed using a Cox proportional hazard model stratified by transitions, given all other risk factors as mentioned above. We restricted our analyses to the impact of ethnicity on transition 1 and 2 i.e. transition from absence of the considered risk factor, and transition from the analysed risk factor to record of stroke respectively.
Long term conditions, socio-demographic variables, and other variables, were displayed as frequencies or median, and interquartile range as appropriate. Cumulative incidence of cardiometabolic stroke risk factors were estimated using the Kalbfleisch and Prentice method [14] versus death and stroke.
All computations were carried out using the R language and environment for statistical computing (version 4·3.0 (2023-04-21))[15].

Ethics

This study was conducted in accordance with the Declaration of Helsinki. Data were provided by the Lambeth DataNet upon approval for the analysis of fully anonymised data by the Lambeth Clinical Commissioning Group. All patients were informed through 'Fair Processing Notices' of the potential use of collected electronic health records in 'secondary data analysis' and were given the option to opt out of data sharing scheme [16]. Accordingly, this study was exempt from ethical committee approval and individual consent requirements as per the National Health Institute (NHI) Health Research Authority (HRA) guidelines for research using anonymised primary care data in the United Kingdom.

Results

Patient characteristics

Of 849,968 registered patients, 651,888 (76·7%) self-identified as being of Asian (8·6%), Black (18·2%), or White (73·2%) ethnicity and 54·1% were female. Patients of White and Asian ethnicity had similar IMD profiles whereas patients of Black ethnicity were overrepresented in IMD quintiles 1 and 2 (most deprived) and conversely less represented in quintiles 4 and 5 (least deprived). Patients of Black ethnicity had the highest rate of incident stroke (2·1%) as well as risk factors including heart failure (1·3%), hypertension (19·7%), hypercholesterolaemia (30·6%), peripheral aterial/vascular disease (0·5%), and diabetes (10·4%). On the other hand, patients of White ethnicity had the lowest rate of incident stroke (1·0%) as well as coronary heart disease (1·2%), heart failure (0·7%), hypertension (6·6%), hypercholesterolaemia (19·0%), and diabetes (2·8%). Finally, patients of Asian ethnicity had intermediate rates of stroke (1·2%), hypertension (10·0%), hypercholesterolaemia (24·7%), and diabetes (8·5%), while having the highest rates of coronary heart disease (2·1%) and myocardial infraction (1·0%) (table 1). Figure 2 displays

the cumulative incidence of hypertension, diabetes, and hypercholesterolaemia (transitions 1) across ethnicity from 18 to 90 years. The populations of Black, Asian, and White ethnicities had, respectively, higher, intermediate, and lower cumulative incidences of hypertension. Both Asian and Black ethnicities were associated with a higher cumulative incidence of diabetes compared to White ethnicity. Finally, while the cumulative incidence of hypercholesterolemia showed less variation between ethnicities, it was notably higher in the Asian population, particularly between the ages of 30 and 50.

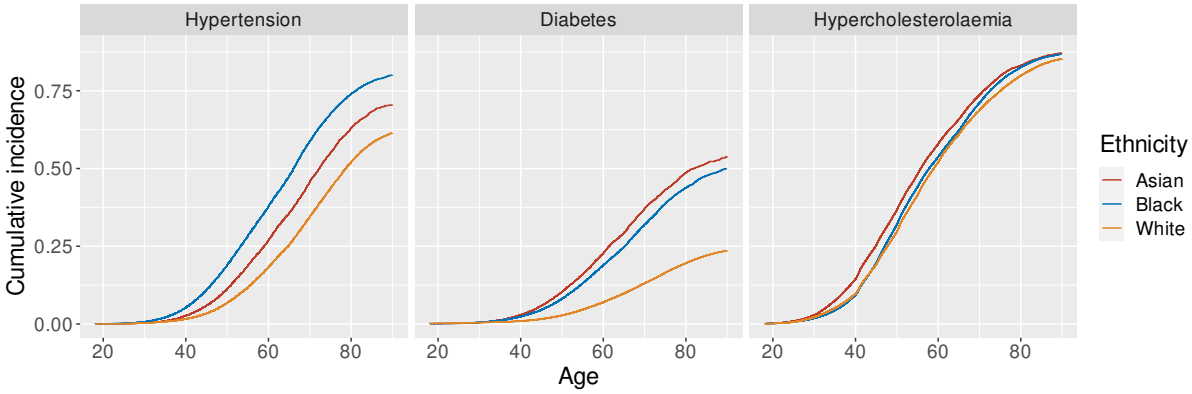

Figure 2: *Cumulative incidence of hypertension, diabetes and hypercholesterolaemia across ethnicity between 18 and 90 years*

Multistate model

Figure 3 illustrates adjusted hazard ratios of Black and Asian ethnicity versus White ethnicity for transitions to hypertension, diabetes, and hypercholesterolaemia (transitions 1), and transitions form these cardiometabolic risk factors to stroke (transitions 2). Ethnicity was significantly associated with transitions to all three risks factor. The highest values were associated with diabetes for Asian and Black ethnicity (HR: 3·04, 95% confidence interval (CI): from 2·94 to 3·15, p<0·001, and HR: 2·33, CI: from 2·27 to 2·40, p<0·001, respectively). Results were more contrasted for hypertension with high to moderate HRs associated with Black and Asian ethnicity respectively (HR: 2·06, CI: from 2·02 to 2·09, p<0·001, and HR: 1·27, CI: from 1·23 to 1·31, p<0·001, respectively). Although highly significant, HRs associated with hypercholesterolaemia were relatively moderate (HR: 1·21, CI: from 1·19 to 1·24, p<0·001, for Asian ethnicity and HR: 1·04, CI: from 1·03 to 1·05, p<0·001 for Black ethnicity). Except for Black ethnicity associated with hypercholesterolaemia (HR: 1·19, CI: from 1·11 to 1·28, p<0·001), ethnicity did not significantly affect hazard of transitions from tested cardiometabolic risk factors to stroke.

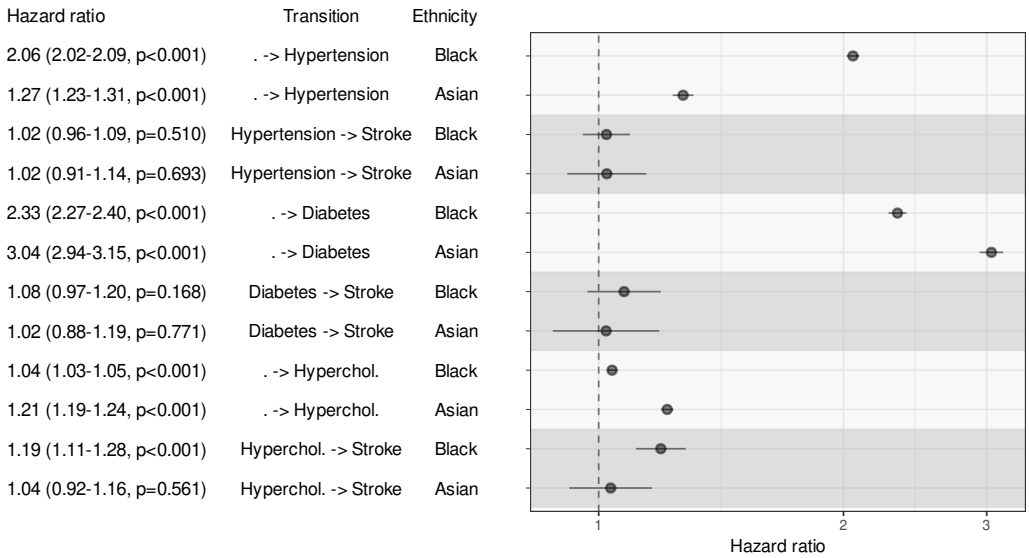

Figure 3: *Transition-specific hazard ratios of the risk factor/disease models in which hypertension, diabetes and hypercholesterolaemia were sequentially analysed as risk factor associated with stroke*

Discussion

This study set out to better understand the impact of ethnicity on the hazard of modifiable cardiometabolic risk factors as well as on the hazard of stroke mediated by these risk factors in a multiethnic and urban population. Our findings do not support the hypothesis of excess risk of stroke in patients of Black ethnicity with hypertension or diabetes. Similar observations were made in patients of Asian ethnicity, not only for hypertension and diabetes but also hypercholesterolaemia. Our results also suggest that patients of Black ethnicity with hypercholesterolemia are at a higher risk of stroke compared to their counterparts of Asian or White ethnicity.

Although previous studies have reported an absence of excess risk of stroke in patients of Black versus White ethnicity with hypertension or diabetes [1, 6], no real consensus has emerged in the literature so far [3]. In line with previous findings, our results support the hypothesis of equipotency of hypertension and diabetes regarding stroke risk in Black versus White Ethnicity, and brings new evidence that this hypothesis also holds for patients of Asian ethnicity.

Our study also confirm the role of hypertension and diabetes as major mediators of the excess risk of stroke in patients of Asian and Black ethnicity [1, 3]. This excess risk of stroke is attributable to the higher incidence and prevalence of these cardiometabolic risk factors in patients of Asian and Black ethnicity (figure 2 and table 1). On the other hand, hypercholesterolaemia, which is more evenly distributed across ethnicity as compared to hypertension and diabetes (table 1, figure 2 and 3), presents overall higher incidence rates (table 1 and figure 2). This establishes hypercholesterolaemia as a major mediator of stroke risk, irrespective of ethnicity [17, 18].

Our findings demonstrate the importance of hypercholesterolaemia in mediating excess risks of

stroke in patients of Asian and Black ethnicity in two different ways. In patients of Asian ethnicity, the excess risk of stroke is directly mediated by the excess risk of hypercholesterolaemia, mainly at an the early age (between 30 and 50 years) (figure 2). In patients of Black ethnicity on the other hand, the excess risk of stroke is mainly due to a higher risk of stroke in patients with hypercholesterolaemia (figure 3).

Of note, the excess incidence of stroke mediated by modifiable cardiometabolic risk factors depends not only on the hazard of stroke associated with these risk factors, but also on their prevalence. Therefore, further investigation, beyond the scope of this study, is needed to better understand ethnic inequality in the incidence of stroke through: i) evaluating the impact of observed ethnic disparity in the hazard of cardiometabolic risk factors (direct mediation), and ii) evaluating the impact of excess risk of stroke associated with ethnicity given cardiometabolic risk factors, such as with hypercholesterolaemia in our study (indirect mediation).
The absence of a significant ethnic-related excess of stroke regarding hypertension and diabetes in our study support the hypothesis of equal adherence and medical care between patients of Asian, Black and White ethnic backgrounds [1]. Though, this question lacks a clear consensus [1, 19], our result indicates that medical care and patient adherence to treatment does not explain the excess risk of stroke associated with hypercholesterolaemia in patients of Black ethnicity.
Other factors, not explored in our study, may bring further insight into the excess risk of stroke observed in patients of Black versus Asian and White ethnicity. These include for instance, levels and ratio of high- and low- density lipoprotein (HDL and LDL respectively) [20], residual confounding associated with differential patterns in socioeconomic status not fully captured by the IMD [6], the impact of mental health conditions [21], the body mass index [22], history of infectious diseases [23] or more generally the patterns of multimorbidity across different ethnic background [24, 25].

Strengths and limitations of our study relate to both the use of routinely collected primary care electronic health records, and the population under study. Though primary care electronic health records are considered a reliable source of data and have been increasingly used in research, surveillance and public health planning [26], concerns may arise regarding accuracy or completeness of collected electronic health records [27]. For instance, information on current smoking status or the level of HDL and LDL was not available in electronic health records analysed in this study.
On the other hand, analysis of longitudinal primary care electronic health records at the population scale enables the conduct of exploratory analysis allowing to characterise moderate differential trends such as the subtle ethnic inequality in the incidence of hypercholesterolaemia reported in our study. This approach offers insights into public health and epidemiology that may be more challenging to obtain through classical approaches, such as case-control studies

or cohort studies [28, 29].
The characteristics of the analysed population are also important aspects to consider here, as they may hinder or favour particular study objectives. For instance, the relative socio-demographic uniformity within the Lambeth population tends to hinder the evaluation of the impact of the socio-economic status on cardiometabolic risk factors and stroke. However, this characteristic is not of prime importance in our study as the socio-economic status is regarded as a confounder. Likewise, the geographical unity of the studied population is also likely to reduce the impact of differential environmental exposure and resulting unidentifiable confounding. Finally, also in line with our objectives, the multiethnic background of the Lambeth population, especially the significant representation of patients of Asian and Black ethnicity, represents an important aspect of our study.

Conclusion

Previous recommendations aiming to reduce inequality in stroke incidence between patients of Black and White ethnicity have focused on reducing ethnic disparities in the incidence of hypertension and diabetes [1, 3], which are major mediators of excess stroke risk in patients of Black ethnicity. Our findings provide new evidence to extend this recommendation to patients of Asian ethnicity as well as for hypercholesterolaemia in these patients. Further research and alternative strategies will be needed to elucidate and address the excess risk of stroke observed in patients of Black ethnicity with hypercholesterolaemia.

Author contribution

MD designed the study (conceptualization and methodology), AD and MA acquired the data, MD carried out data curation, MD and MA conducted investigations, MD provided computer code and performed formal analysis including data visualisation and validation, MD, MA and AD interpreted the data. MD drafted the manuscript and AD supervised the research. MA, MA and AD reviewed, edited and approved the final version of the manuscript.

Data availability

Legal restrictions apply to the availability of the data that support the findings of this study, as they were used under an IRB approval and are hence not publicly available.

Acknowledgement

This project is funded by Guys and St Thomas Charity MLTC Challenge Fund (grant number EIC180702). The founder had no role in the study design, data collection, analysis, interpretation, or writing of the report, nor in the decision to submit the paper for publication.

Table 1: *Sociodemographics and of stroke risk factors across ethnic groups*

Ethnicity		Asian			Black			White			Total
		No stroke	Stroke	Sub-total	No stroke	Stroke	Sub-total	No stroke	Stroke	Sub-total	
Total N (%)		55417 (98·8)	645 (1·2)	56062 (8·6)	116439 (97·9)	2495 (2·1)	118934 (18·24)	471893 (99·0)	4999 (1·0)	476892 (73·16)	651888
Age at end of follow-up	Median (IQR)	35·2 (18·3)	76·1 (20·1)	35·5 (18·6)	42·1 (24·0)	72·1 (23·0)	42·1 (24·8)	34·1 (15·0)	75·9 (24·0)	34·1 (15·0)	35·1 (17·9)
Sex	Male	25459 (45·9)	368 (57·1)	25827 (46·1)	54626 (46·9)	1310 (52·5)	55936 (47·0)	214811 (45·5)	2695 (53·9)	217506 (45·6)	299269 (45·9)
	Female	29958 (54·1)	277 (42·9)	30235 (53·9)	61813 (53·1)	1185 (47·5)	62998 (53·0)	257067 (54·5)	2304 (46·1)	259371 (54·4)	352604 (54·1)
IMD[a]	1 (most deprived)	7295 (13·4)	105 (16·4)	7400 (13·5)	33957 (29·5)	758 (30·6)	34715 (29·5)	64970 (14·0)	1117 (22·5)	66087 (14·0)	108202 (16·8)
	2	24319 (44·8)	289 (45·2)	24608 (44·8)	56405 (48·9)	1169 (47·1)	57574 (48·9)	217833 (46·8)	2112 (42·6)	219945 (46·7)	302127 (47·0)
	3	15377 (28·3)	178 (27·9)	15555 (28·3)	20833 (18·1)	467 (18·8)	21300 (18·1)	136289 (29·3)	1347 (27·2)	137636 (29·2)	174491 (27·1)
	4	6596 (12·1)	60 (9·4)	6656 (12·1)	3636 (3·2)	78 (3·1)	3714 (3·2)	36634 (7·9)	327 (6·6)	36961 (7·9)	47331 (7·4)
	5 (least deprived)	752 (1·4)	7 (1·1)	759 (1·4)	407 (0·4)	9 (0·4)	416 (0·4)	9936 (2·1)	55 (1·1)	9991 (2·1)	11166 (1·7)
Hypertension		5220 (9·4)	366 (56·7)	5586 (10·0)	21990 (18·9)	1498 (60·0)	23488 (19·7)	29195 (6·2)	2167 (43·3)	31362 (6·6)	60436 (9·3)
Diabetes		4519 (8·2)	231 (35·8)	4750 (8·5)	11680 (10·0)	735 (29·5)	12415 (10·4)	12574 (2·7)	687 (13·7)	13261 (2·8)	30426 (4·7)
Hypercholesterolaemia		13545 (24·4)	326 (50·5)	13871 (24·7)	35208 (30·2)	1167 (46·8)	36375 (30·6)	88094 (18·7)	2299 (46·0)	90393 (19·0)	140639 (21·6)
Coronary Heart Disease		1044 (1·9)	127 (19·7)	1171 (2·1)	1375 (1·2)	192 (7·7)	1567 (1·3)	5327 (1·1)	612 (12·2)	5939 (1·2)	8677 (1·3)
Heart Failure		406 (0·7)	39 (6·0)	445 (0·8)	1340 (1·2)	151 (6·1)	1491 (1·3)	3186 (0·7)	240 (4·8)	3426 (0·7)	5362 (0·8)
PA/VD[b]		168 (0·3)	26 (4·0)	194 (0·3)	505 (0·4)	66 (2·6)	571 (0·5)	1879 (0·4)	200 (4·0)	2079 (0·4)	2844 (0·4)
Myocardial infarction		493 (0·9)	54 (8·4)	547 (1·0)	624 (0·5)	92 (3·7)	716 (0·6)	2829 (0·6)	345 (6·9)	3174 (0·7)	4437 (0·7)
Atrial Fibrillation		295 (0·5)	55 (8·5)	350 (0·6)	896 (0·8)	122 (4·9)	1018 (0·9)	5041 (1·1)	570 (11·4)	5611 (1·2)	6979 (1·1)
Ever Smoked		16413 (29·6)	204 (31·6)	16617 (29·6)	40267 (34·6)	991 (39·7)	41258 (34·7)	227286 (48·2)	2656 (53·1)	229942 (48·2)	287817 (44·2)

[a] Index of multiple deprivation

[b] Peripheral Aterial/Vascular Disease

References

[1] Steven J Kittner, Lon R White, Katalin G Losonczy, Philip A Wolf, and J Richard Hebel. Black-white differences in stroke incidence in a national sample: the contribution of hypertension and diabetes mellitus. *Jama*, 264(10):1267–1270, 1990.

[2] Judith A Stewart, R Dundas, RS Howard, AG Rudd, and CDA Wolfe. Ethnic differences in incidence of stroke: prospective study with stroke register. *Bmj*, 318(7189):967–971, 1999.

[3] George Howard, Mary Cushman, Brett M Kissela, Dawn O Kleindorfer, Leslie A McClure, Monika M Safford, J David Rhodes, Elsayed Z Soliman, Claudia S Moy, Suzanne E Judd, et al. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. *Stroke*, 42(12):3369–3375, 2011.

[4] Wilson Nadruz Jr, Brian Claggett, Mir Henglin, Amil M Shah, Hicham Skali, Wayne D Rosamond, Aaron R Folsom, Scott D Solomon, and Susan Cheng. Racial disparities in risks of stroke. *New England Journal of Medicine*, 376(21):2089–2090, 2017.

[5] Hannah Gardener, Ralph L Sacco, Tatjana Rundek, Valeria Battistella, Ying Kuen Cheung, and Mitchell SV Elkind. Race and ethnic disparities in stroke incidence in the northern manhattan study. *Stroke*, 51(4):1064–1069, 2020.

[6] Dawn M Bravata, Carolyn K Wells, Barbara Gulanski, Walter N Kernan, Lawrence M Brass, Judith Long, and John Concato. Racial disparities in stroke risk factors. 2005.

[7] M Maria Glymour, Mauricio Avendaño, Steven Haas, and Lisa F Berkman. Lifecourse social conditions and racial disparities in incidence of first stroke. *Annals of epidemiology*, 18(12):904–912, 2008.

[8] Nicole D Armstrong, Vinodh Srinivasasainagendra, Amit Patki, Rikki M Tanner, Bertha A Hidalgo, Hemant K Tiwari, Nita A Limdi, Ethan M Lange, Leslie A Lange, Donna K Arnett, et al. Genetic contributors of incident stroke in 10,700 african americans with hypertension: A meta-analysis from the genetics of hypertension associated treatments and reasons for geographic and racial differences in stroke studies. *Frontiers in Genetics*, 12:781451, 2021.

[9] Matthew Traylor, Loes Rutten-Jacobs, Charles Curtis, Hamel Patel, Gerome Breen, Stephen Newhouse, Cathryn M Lewis, and Hugh S Markus. Genetics of stroke in a uk african ancestry case-control study: South london ethnicity and stroke study. *Neurology Genetics*, 3(2), 2017.

[10] Kahlilia C Morris-Blanco, Anil K Chokkalla, Vijay Arruri, Soomin Jeong, Samantha M Probelsky, and Raghu Vemuganti. Epigenetic mechanisms and potential therapeutic targets in stroke. *Journal of Cerebral Blood Flow & Metabolism*, 42(11):2000–2016, 2022.

[11] Souvik Sen, Katherine Dahlberg, Abigail Case, Stephanie Paolini, Joselyn Burdine, Leema R Peddareddygari, and Raji P Grewal. Racial-ethnic differences in stroke risk factors and subtypes: results of a prospective hospital-based registry. *International Journal of Neuroscience*, 123(8):568–574, 2013.

[12] Sophie V Eastwood, Therese Tillin, Nish Chaturvedi, and Alun D Hughes. Ethnic differences in associations between blood pressure and stroke in south asian and european men. *Hypertension*, 66(3):481–488, 2015.

[13] Richard J Cook and Jerald F Lawless. *Multistate models for the analysis of life history data*. CRC Press, 2018.

[14] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*. John Wiley & Sons, 2011.

[15] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.

[16] NHS England. Privacy Notice - NHS England and NHS Improvement. Accessed on 2024-04-01.

[17] Martin J O'donnell, Denis Xavier, Lisheng Liu, Hongye Zhang, Siu Lim Chin, Purnima Rao-Melacini, Sumathy Rangarajan, Shofiqul Islam, Prem Pais, Matthew J McQueen, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the interstroke study): a case-control study. *The Lancet*, 376(9735):112–123, 2010.

[18] Farshad Farzadfar, Mariel M Finucane, Goodarz Danaei, Pamela M Pelizzari, Melanie J Cowan, Christopher J Paciorek, Gitanjali M Singh, John K Lin, Gretchen A Stevens, Leanne M Riley, et al. National, regional, and global trends in serum total cholesterol since 1980: systematic analysis of health examination surveys and epidemiological studies with 321 country-years and 3· 0 million participants. *The Lancet*, 377(9765):578–586, 2011.

[19] A Agarwala, N Bekele, E Deych, MW Rich, A Hussain, LK Jones, et al. Racial disparities in modifiable risk factors and statin usage in black patients with familial hypercholesterolemia. j am heart assoc. 2021; 10 (17): e020890.

[20] Samantha A Reina, Maria M Llabre, Matthew A Allison, John T Wilkins, Armando J Mendez, Martinson K Arnan, Neil Schneiderman, Ralph L Sacco, Mercedes Carnethon, and JA Chris Delaney. Hdl cholesterol and stroke risk: The multi-ethnic study of atherosclerosis. *Atherosclerosis*, 243(1):314–319, 2015.

[21] Jia-Yi Dong, Yong-Hong Zhang, Jian Tong, and Li-Qiang Qin. Depression and risk of stroke: a meta-analysis of prospective studies. *Stroke*, 43(1):32–37, 2012.

[22] Tobias Kurth, J Michael Gaziano, Klaus Berger, Carlos S Kase, Kathryn M Rexrode, Nancy R Cook, Julie E Buring, and JoAnn E Manson. Body mass index and the risk of stroke in men. *Archives of internal medicine*, 162(22):2557–2562, 2002.

[23] Jennifer E Fugate, Jennifer L Lyons, Kiran T Thakur, Bryan R Smith, E Tessa Hedley-Whyte, and Farrah J Mateen. Infectious causes of stroke. *The Lancet Infectious Diseases*, 14(9):869–880, 2014.

[24] Katie I Gallacher, Bhautesh D Jani, Peter Hanlon, Barbara I Nicholl, and Frances S Mair. Multimorbidity in stroke. *Stroke*, 50(7):1919–1926, 2019.

[25] Marc Delord, Xiaohui Sun, Annastazia Learoyd, Vasa Curcin, Iain Marshall, Charles Wolfe, Mark Ashworth, and Abdel Douiri. Patient-oriented unsupervised learning to unlock patterns of multimorbidity associated with stroke using primary care electronic health records. *arXiv preprint arXiv:2401.01870*, 2024.

[26] George Hripcsak and David J Albers. Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association*, 20(1):117–121, 2013.

[27] Nicole Gray Weiskopf and Chunhua Weng. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, 20(1):144–151, 2013.

[28] Carol C Diamond, Farzad Mostashari, and Clay Shirky. Collecting and sharing data for population health: a new paradigm. *Health affairs*, 28(2):454–466, 2009.

[29] Clemens Scott Kruse, Anna Stein, Heather Thomas, and Harmander Kaur. The use of electronic health records to support population health: a systematic review of the literature. *Journal of medical systems*, 42(11):214, 2018.