BMJ Open

Diagnosis of cerebrovascular disease within the UK Biobank; characteristics of 502650 participants.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-009161
Article Type:	Cohort profile
Date Submitted by the Author:	29-Jun-2015
Complete List of Authors:	Hewitt, Jonathan; Cardiff University, Geriatric Medicine Walters, Matthew; University of Glasgow, Padmanabhan, Sandosh; University of Glasgow, Dawson, Jesse; University of Glasgow,
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Geriatric medicine
Keywords:	Stroke < NEUROLOGY, EPIDEMIOLOGY, GERIATRIC MEDICINE

Diagnosis of cerebrovascular disease within the UK Biobank; characteristics of 502650 participants.

Hewitt J[1], Walters M[2], Padmanabhan S[2] and Dawson J[2]

Jonathan Hewitt, MBBS, MSc, FRCP, PhD, Clinical Senior Lecturer and Corresponding Author

Mathew Walters, MD, MBChB, FRCP, Professor

Sandosh Padmanabhan, MD, PhD, FRCP, FAHA, Professor

Jesse Dawson, MD, MBChB (Hons), FRCP Clinical Reader

1. Department of Geriatric Medicine, Institute of Primary Care and Public Health, Cardiff University.

Address: Academic Building, University Hospital Llandough, Cardiff, CF64 2XX, Tel 02920 716982, Fax 02920 704244, Email hewittj2@cardiff.ac.uk, Corresponding author.

2. Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow.

Word Count 2994

Classification and description of Stroke Disease using the UK Biobank

Abstract

Objectives and setting

The UK Biobank is a large-scale biomedical resource, containing socio demographic and medical information, including data on a previous diagnosis of stroke or transient ischaemic attack (TIA). We described these participants and their medication usage.

Participants and outcomes

We identified participants who either self reported or were identified from a nurse led interview, having suffered a stroke or a transient ischemic attack (TIA). We assessed their risk factor burden (sex, age, deprivation, waist to hip ratio (WHR), hypertension, smoking, alcohol intake, diabetes, physical exercise and oral contraception use (OCP)) and medication usage.

Results

We studied 502,650 people (54.41% women), 6669 (1.23%) participants self reported a stroke. The nurse led interview identified 7669 (1.53%) people and 1781 (0.35%) with TIA. Hypertension, smoking, higher WHR, lower alcohol consumption and diabetes were all more common in people with cerebrovascular disease ($p<0.0001$ for each). Women were less likely to have taken the OCP ($p=0.0002$). People with stroke disease did more exercise ($p=0.03$). Antithrombotic medication was taken by 81% of people with stroke (both self report and nurse led responders) and 89% with TIA. For self reported stroke, 63% were taking antithrombotic and cholesterol medications, 54% taking antithrombotic and anti-hypertensive medications and 46% taking all three. For the nurse led interview and TIA these figures were 65%, 54% and 46% and 70%, 53% and 45% respectively.

Conclusions

The UK Biobank provides a large, generalisable and contemporary data source in a young population. Our findings represent the initial characterisation of the UK Biobank participants with cerebrovascular disease.

Article Summary

Strengths and Weaknesses

A very large sample

Self Reported data

Key Words

Stroke, UK Biobank, Prevalence, Risk Factors

Introduction

The UK Biobank is a national health resource, intended to improve the prevention, diagnosis and treatment of illness [1]. It holds detailed demographic, social and medical information, along with physical measures, such as weight and blood pressure. Participants gave consent to have their future health records linked to their UK Biobank data, a process which is on going.

There is comparatively little epidemiological data on younger populations with stroke[2]. The UK Biobank obtained information regarding previous stroke or transient ischaemic attack and commonly associated medications (such as blood pressure and lipid lowering medications) in people aged 40-69 years old. It is an order of magnitude larger than previous self reported estimates of cerebrovascular disease. We aimed to describe the participants from the UK Biobank who self reported a diagnosis of stroke and transient ischaemic attack.

We hypothesised that the data within the UK Biobank, will be comparable to other cohorts of self reported cerebrovascular data and that this population will not be receiving optimal medication.

Methods

Between 2006 and 2010, the UK Biobank collected detailed data on 500000 people. They used 22 assessment centres based in England, Scotland and Wales. Participants were recruited from NHS patient registers and contacted if they lived within a reasonable proximity to an assessment centre [3, 4].

Participants provided detailed demographic, socioeconomic and health related data via a touchscreen questionnaire, including medication history. One of the touchscreen questionnaire responses was “Has a doctor ever told you that you have any of the following

conditions? (You can select more than one answer)" The possible touch screen responses were "Heart Attack, Angina, Stroke, High blood pressure, None of the above, Prefer not to answer".

Following the questionnaire, a nurse led interview was performed to address past medical history in general. In order to prompt the nurses conducting the questionnaire and improve accuracy, all responses from the touch screen questionnaire were flagged to the study nurse. Therefore if a participant had selected "stroke" from the touchscreen question listed above then the nurse would have been aware of this response when discussing past medical history. The nurses also asked individuals directly about their past medical history. Therefore it is possible that someone may not have responded positively to the touchscreen questionnaire question regarding stroke but may have done so when asked later by the study nurse about past medical history. It is also during the nurse led questionnaire that participants would have been identified as have a past medical history of TIA.

Participants were also asked to self-report via the touchscreen questionnaire their medication history. They were asked specifically if they regularly took "Medication for cholesterol" or "blood pressure medication" and could answer positively for both responses,

During the nurse led interview all medications where recorded as free text and subsequently grouped and coded by UK Biobank. We extracted data on reported antiplatelet or anticoagulant use. Any participant using any single or combination of these drugs was categorised as taking antithrombotic medication.

We also extracted demographic data concerning age, sex and socio-economic class, measured using the Townsend scale[5]. In order to select the most appropriate potential risk factors for stroke to study, we used data from the INTERSTROKE study [6] and the American Heart Associations' Stroke Councils Scientific Statement Oversight Committee guideline (AHA)[7].

Review of these publications identified seven modifiable risk factors that we could accurately be reproduced from the UK Biobank. These risk factors represent the majority of the modifiable risk factors contributing to cerebrovascular disease.

From the INTERSTROKE study we identified hypertension, smoking, waist to hip ratio, regular physical activity, diabetes mellitus and alcohol intake as risk factors we could study. We could not include psychological factors, diet, cardiac causes and apolipoprotein levels as the data were not available in comparable format. From the AHA guideline we also identified the oral contraceptive pill (OCP). We used the same definitions as INTERSTROKE and the AHA guideline as far as possible; we defined a self-reported history of hypertension, smoking (current, ex or never), waist to hip ratio (divided into tertiles), moderate or strenuous activity for more than 4 hours per week and a self reported history of diabetes mellitus. For women, a history of ever having taken the oral contraceptive pill (yes or no).

Alcohol intake was based on the data recorded in the UK Biobank (alcohol daily or almost daily, alcohol on three or four days per week, alcohol once or twice a week, alcohol one to three times per month, special occasions only, never drinkers and former drinkers).

We describe the characteristics of the population using descriptive statistics. We compared characteristics of patients who reported stroke and TIA with those who did not using t tests for continuous variables and chi-squared tests for dichotomous variables. The statistical analysis was conducted using STATA version 11 (Stata Corp).

Results

UK Biobank released data for use on 502650 participants. Their average age was 56.52 years, (standard deviation (SD) 8.09). There were 273468 women (54.41%) who were younger than

the men (56.35 years (SD 8.00) compared to 56.74 years (SD 8.20), p<0.0001. The mean Townsend Index for all participants was -1.29 (SD 3.10).

More participants self-reported stroke during the nurse led interview (n=7669 (1.53%)) than on the touchscreen questionnaire (n=6699 (1.33%)). TIA was reported via the verbal questionnaire in 1781 (0.35%) of participants. Patients self-reporting stroke and TIA were older and more likely to be female (p<0.0001) in comparison to patients who did not. Higher levels of deprivation were seen in people self reporting stroke disease via both the touchscreen and nurse led questionnaire (p<0.0001) but this was not the case for reports of TIA (p=0.13). The full results are shown in table 1.

The cardiovascular risk factors in patients self-reporting stroke and TIA are shown in table 2 along with patients who reported no stroke or TIA.

Self-reported hypertension was recorded in 27.01% of the whole UK Biobank population. It was over 50% for each of the patients with a history of self reported stroke, stroke from the nurse led questionnaire and TIA. People with any cerebrovascular history were more likely to be former or current smokers. Diabetes was reported in 5.25% of the UK Biobank population but over 15% of stroke patients and 10% of patients with TIA. Each of these associations was significant (p<0.001, for each).

Self-reported alcohol intake was lower in people with cerebrovascular disease in comparison to the population without cerebrovascular disease. In people who drank less than weekly, including never and former drinkers, there was a higher proportion of cerebrovascular disease recorded (p<0.001) compared to those who drank more frequently.

For women, there were 81.05% who had previously taken the oral contraceptive pill. For self-reported stroke and nurse led stroke responses, these figures were 73.83% and 74.22% respectively (p<0.001) and for TIA, this figure was 75.55% (p=0.0002).

There were 23823 (4.74%) people who reported doing at least 4 or more hours of moderate or strenuous physical exercise per week. Of the 303897 people with complete data for this response 302 (6.4%) of those who had a previous stroke identified via the touchscreen questionnaire responded positively to this question (versus 17458 (5.84%), p=0.1, for those without a previous stroke). For the nurse led self-report question these figures were 357 (6.47%) and 23463 (5.78%), p=0.03 and self-reported TIA they were 91 (6.45%) versus 17669 (5.84%), p=0.32.

In total, 76397 (15.20%) people who were identified as taking an antithrombotic medication, 104028 (20.70%) of people self-reported as receiving a blood pressure medication and 86907 (17.29%) of people self-reported receiving a cholesterol medication. For each of these responses participants who reported stroke or TIA disease all received more of these medications than the population as a whole, every response was significant (<0.0001). The full results are given in table 3.

Discussion

The objective of this study was to characterise the UK Biobank participants with a history of stroke or TIA and to gain insights into the treatment of patients at population level in the UK.

There are surprisingly few studies documenting the prevalence of stroke and TIA disease. For example, a 2012 review [2] identified only five studies of stroke prevalence in the Western world. They estimated that for an adult population, including older populations, the prevalence of stroke ranged from 0.15% in Italy to between 1.7% and 2.6% in the UK and the US. When these figures were assessed with respect to sex, there was a suggestion that

prevalence was higher in men. Age is by far the largest non-modifiable risk factor for stroke making the findings in our younger cohort (1.74% for men and 0.99% for women) comparable.

Our findings noted a difference in reporting of prevalence between the self reported and nurse led figures for stroke disease. There were 970 more people identified with stroke disease following the nurse led questionnaire. Every person who responded positively on the touch screen questionnaire to having stroke disease will have been challenged by the nurse during the nurse led interview. All of these people were included in the nurse led responses as having a stroke, therefore this data has been at least partially verified by the nurse. The nurse also further identified people with stroke disease via questions on past medical history. We would advocate that the figure (7669) is the most accurate estimation of stoke disease within the UK Biobank.

Reduced socio-economic status was associated with both self and nurse reported stroke and TIA. Deprivation and stroke are associated with an increase in mortality across a range of countries[8] and also a Scottish meta-analysis linked deprivation with stroke incidence even after allowing for cardiovascular risk factors[9]. However, our data is an order of magnitude larger than the studies included in that review and adds a contemporary UK wide assessment of stroke and deprivation.

When selecting the risk factors to study from the UK Biobank we attempted to use the risk factors characterised in other studies. The advantages of this approach are that the reproducibility of our study is increased and our results can be compared directly to those generated from the high-income countries included in their studies. We were able to reproduce 6 of the risk factors used in INTERSTROKE study and the AHA guideline.

Hypertension was reported by 27% of the whole cohort but in over 50% of the participants with cerebrovascular disease. These figures are consistent with other epidemiological studies[10]. Our findings reflect both the causative effect of blood pressure and a heightened awareness of hypertension treatment in stroke survivors. Unsurprisingly our results show an increased number of current smokers and former smokers and a reduced rate of never smokers in those reporting cerebrovascular disease.

UK Biobank waist to hip ratio (WHR) data can be closely compared to the INTERSTROKE study data. This is because we were able to conduct our analysis using tertiles and also because while weight may have altered following a diagnosis of cerebrovascular disease it is likely to be one of the most challenging of modifiable risk factors to change following diagnosis, although data on post stroke weight change is sparse[11]. In the INTERSTROKE study, the incident rates were 23%, 33% and 43% for the increasing tertiles of WHR. These results are comparable to our results for both self-reported TIA (22%, 33% and 45% respectively) and stroke (18%, 30% and 51%). The high figure of 51% implies either an increased weight gain following stroke or a lower prevalence of obesity in the INTERSTROKE study, which included data from both high and low socioeconomic countries.

Our results suggest that people who undertook at least four hours of physical activity per week were more likely to have had a cerebrovascular event. Four hours is a relatively large amount and studies have consistently shown benefit in increased physical activity in reducing stroke disease[12]. Our results are likely to be due to reverse causation; either increased activity in cerebrovascular sufferers to improve their general health or the residual physical disability being reflected in normal activities of daily living being deemed at least moderately strenuous.

The self-reported history of diabetes mellitus was 5.25%. For a cohort of middle aged, predominantly Caucasian Europeans, this is what we would have anticipated[13] . Estimates put diabetes prevalence between 10 and 25% in people with existing cerebrovascular disease depending on population and method used to diagnose diabetes[14] and our results (15% for stroke and 10% for TIA) compare closely to a self-reported diagnosis in a young population.

In people with cerebrovascular disease who drank 3 times or more per month, their frequency of drinking was below people without cerebrovascular disease. Further, in people who have never drunk or no longer drink there was more cerebrovascular disease, which may reflect either a reduced frequency of drinking following diagnosis or an increased prevalence of disease; the well described "J shaped curve" of alcohol related illness[15].

The oral contraceptive pill (OCP) is associated with ischaemic stroke, particularly at higher doses[16, 17]. We found that use of the oral contraceptive was lower for our cerebrovascular outcomes. Hence, it is possible that some of our participants with a past medical history of stroke disease were not prescribed the OCP, explaining the findings.

A high proportion of participants with self reported and nurse reported stroke (81%) and TIA (89%) were receiving an antithrombotic medication. The figure of 81% would seem appropriate for stroke, because up to 20% of people may have suffered intracerebral haemorrhage, thus restricting use and antithrombotic medication is not suitable for every individual. Our findings compare favourably to the Post Stroke Rehabilitation Outcome Project[18]. Their study in 2005 of 1161 people with both ischemic and haemorrhagic stroke, identified 62% of people as receiving antithrombotic medication.

While stroke disease does appear to have appropriate rates of antithrombotic prescription, TIA does appear to be under treated for antithrombotic medication. TIA is almost exclusively a disease results from infarction, embolic or otherwise, rather than haemorrhage. While

antithrombotic will not be suitable for all of these people it is likely that 11% of untreated TIA disease represents suboptimal management.

We also demonstrated that large numbers of positive responders were taking cholesterol and blood pressure medications, including combinations of these medications. Not all of these medications will be suitable for each person with cerebrovascular disease, however, only about 45% of participants were receiving all three medications, suggesting these medications are under prescribed overall. Neither did we have information on whether treatment targets had been adequately achieved with these medications, something which is also likely to be suboptimal[19].

It is important to highlight what we have not done in this study. Firstly, we did not consider all of the well recognised risk factors in as much detail as possible, for example, psychological disease. Smith and colleagues recently proposed criteria for probable major depression and probable bipolar disorder within the UK Biobank [20]. Currently these criteria are not directly available from UK Biobank. As these and other criteria become established researchers will be able to test hypotheses directly between them and cerebrovascular disease in the UK Biobank. Secondly we have not looked at many of the known but less well established risk factors. There were two reasons for not doing so. Often this was where the exact question did not exist within the UK Biobank, for example recreational drug use. The second reason was where clinical information was available it would be best studied in relation to future linked data. For example, will the people who self reported an asymptomatic carotid stenosis incur a different rate of incident cerebrovascular disease? Finally and perhaps most importantly this is self reported data. As such it is likely that it will under estimate the true prevalence of self reported and nurse reported stroke disease, as well as TIA.

Despite the limitations of the data, this sample of half a million people still represents the largest self reported estimate of cerebrovascular disease in a population of this or any age. It adds to the available data on young populations with self reported stroke and TIA disease which are less well described than older populations. The clinical relevance of the data is to highlight that many people were not receiving all of the secondary prevention medication that would have been expected.

Contributions, access fees, disclosures and data sharing statements.

Hewitt conceived the project and performed all data analysis. Hewitt and Dawson wrote the first draft. All authors developed the project, commented on results and contributed to all subsequent draft.

The access fee for use of the UK Biobank was provided by the College of Medical Veterinary and Life Sciences, University of Glasgow.

Jonathan Hewitt is the Vice Chairman of the Ethics and Governance committee of the UK Biobank. None of the authors have disclosures and conflict of interest.

Data is available from the UK Biobank for all bone fide researchers.

References

1. Collins R. What makes uk biobank special? *Lancet*. 2012;379:1173-1174
2. Zhang Y, Chapman AM, Plested M, Jackson D, Purroy F. The incidence, prevalence, and mortality of stroke in france, germany, italy, spain, the uk, and the us: A literature review. *Stroke research and treatment*. 2012;2012:436125
3. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, et al. Uk biobank: Current status and what it means for epidemiology. *Health Policy and Technology*. 2012;1:123-126
4. www.ukbiobank.ac.uk.
5. Phillimore P, Beattie A, Townsend P. Widening inequality of health in northern england, 1981-91. *Bmj*. 1994;308:1125-1128
6. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the interstroke study): A case-control study. *Lancet*. 2010;376:112-123
7. Goldstein LB, Adams R, Alberts MJ, Appel LJ, Brass LM, Bushnell CD, et al. Primary prevention of ischemic stroke: A guideline from the american heart association/american stroke association stroke council: Cosponsored by the atherosclerotic peripheral vascular disease interdisciplinary working group; cardiovascular nursing council; clinical cardiology council; nutrition, physical activity, and metabolism council; and the quality of care and outcomes research interdisciplinary working group: The american academy of neurology affirms the value of this guideline. *Stroke; a journal of cerebral circulation*. 2006;37:1583-1633
8. Wu SH, Woo J, Zhang XH. Worldwide socioeconomic status and stroke mortality: An ecological study. *International journal for equity in health*. 2013;12:42
9. Kerr GD, Slavin H, Clark D, Coupar F, Langhorne P, Stott DJ. Do vascular risk factors explain the association between socioeconomic status and stroke incidence: A meta-analysis. *Cerebrovascular diseases*. 2011;31:57-63
10. Kannel WB. Blood pressure as a cardiovascular risk factor: Prevention and treatment. *JAMA : the journal of the American Medical Association*. 1996;275:1571-1576
11. Jonsson AC, Lindgren I, Norrving B, Lindgren A. Weight loss after stroke: A population-based study from the lund stroke register. *Stroke; a journal of cerebral circulation*. 2008;39:918-923
12. Lee CD, Folsom AR, Blair SN. Physical activity and stroke risk: A meta-analysis. *Stroke; a journal of cerebral circulation*. 2003;34:2475-2481
13. Group DS. Age- and sex-specific prevalences of diabetes and impaired glucose regulation in 13 european cohorts. *Diabetes care*. 2003;26:61-69
14. Hewitt J, Castilla Guerra L, Fernandez-Moreno M, Sierra C. Diabetes and stroke prevention: A review. *Stroke research and treatment*. 2012
15. Ferreira MP, Weems MK. Alcohol consumption by aging adults in the united states: Health benefits and detriments. *Journal of the American Dietetic Association*. 2008;108:1668-1676
16. Peragallo Urrutia R, Coeytaux RR, McBroom AJ, Gierisch JM, Havrilesky LJ, Moorman PG, et al. Risk of acute thromboembolic events with oral contraceptive use: A systematic review and meta-analysis. *Obstetrics and gynecology*. 2013;122:380-389
17. Zeitoun K, Carr BR. Is there an increased risk of stroke associated with oral contraceptives? *Drug safety : an international journal of medical toxicology and drug experience*. 1999;20:467-473
18. Zorowitz RD, Smout RJ, Gassaway JA, Horn SD. Antiplatelet and anticoagulant medication usage during stroke rehabilitation: The post-stroke rehabilitation outcomes project (psrop). *Topics in stroke rehabilitation*. 2005;12:11-19
19. McAlister FA, Majumdar SR, Padwal RS, Fradette M, Thompson A, Buck B, et al. Case management for blood pressure and lipid level control after minor stroke: Prevention randomized controlled trial. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2014;186:577-584

20. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within uk biobank: Cross-sectional study of 172,751 participants. *PloS one*. 2013;8:e75362

		Sex		Age	Deprivation
Self Reported Stroke via Touchscreen Questionnaire					
		Men	Women	(Standard Deviation)	(Standard Deviation)
Yes	6699 (1.33)	3983 (1.74)	2716 (0.99)	60.8 (6.78)	-0.25 (3.54)
No	368451 (73.3)	167096 (72.91)	201355 (73.63)	57.4 (6.78)	-1.25 (3.11)
Missing (or not completed)	127506 (25.37)	58103 (25.35)	69403 (25.38)		
Self Reported Stroke via Nurse Interview					
Yes	7669 (1.53)	4516 (1.97)	3153 (1.15)	60.89 (6.73)	-0.42 (3.48)
No	494059 (98.29)	224205 (97.83)	269852 (98.68)	56.4 (8.09)	-1.31 (3.08)
Missing (or not completed)	930 (0.19)	461 (0.2)	469 (0.17)		
Self Reported TIA via Nurse Interview					
Yes	1781 (0.35)	170121 (74.23)	203248 (74.32)	62.19 (5.87)	-1.31 (3.05)
No	373369 (74.28)	958 (0.42)	823 (0.3)	57.19 (7.9)	-1.22 (3.12)
Missing (or not completed)	127506 (25.37)	58103 (25.35)	69403 (25.38)		
Number (%)	502650 (100)	229182 (100)	273468 (100)		

Table 1. Self report for stroke and TIA with demographics

		Total UK Biobank	Stroke Diagnosed by touchscreen questionnaire	Stroke Diagnosed by nurse led questionnaire	Transient Ischemic Attack (nurse led questionnaire)
		N=502650 (100%)	N=6669	N=7669	N=1781
Hypertension (self reported)		135787 (27.06)	3770 (56.34)	4257 (55.51)	912 (51.21)
Smoking	Current	52989 (10.92)	1103 (16.48)	1223 (15.95)	191 (10.72)
	Former	173203 (34.52)	2851 (42.60)	3250 (42.38)	731 (41.04)
	Never	275565 (54.92)	2738 (40.91)	3196 (41.67)	859 (48.23)
Waist to Hip Ratio (tertiles)	Lowest	N/A	1181 (17.63)	1375 (17.93)	399 (22.4)
	Middle	N/A	1987 (29.66)	2308 (30.10)	574 (32.23)
	Highest	N/A	3470 (51.80)	3913 (51.02)	801 (44.97)
Self Report Diabetes	Yes	26408 (5.26)	1042 (15.57)	1151 (15.01)	178 (9.99)
Alcohol frequency	Daily or almost daily	101794 (20.29)	1217 (18.19)	1437 (18.74)	389 (21.84)
	3-4 times/week	115459 (23.01)	1100 (16.44)	1280 (16.69)	340 (19.09)
	1-2/week	129325 (25.77)	1560 (23.31)	1800 (23.47)	406 (22.8)
	upto 3/month	55871 (11.14)	713 (10.65)	821 (10.71)	200 (11.23)
	Special Occasions	58032 (11.57)	1010 (15.09)	1132 (14.76)	233 (13.08)
	Never	22551 (4.49)	476 (7.11)	529 (6.9)	108 (6.06)
	Former	18115 (3.61)	597 (8.92)	651 (8.49)	104 (5.84)
Oral Contraceptive Pill (N=272047)	Yes	220501 (81.05)	1989 (73.83)	2326 (74.22)	621 (75.55) (p=0.002)

Table 2, Descriptive Characteristics (p<0.001, unless stated)

	Total UK Biobank	Stroke Diagnosed by touchscreen Questionnaire	Stroke Diagnosed by nurse led Questionnaire	Transient Ischemic Attack (nurse led questionnaire)
	N=502650 (100%)	N=6669	N=7669	N=1781
Antithrombotic Medications	76397 (15.20)	5414 (80.82)	6243 (81.41)	1577 (88.55)
Cholesterol Medications	86907 (17.29)	4728 (70.58)	5466 (71.27)	1303 (73.16)
Blood Pressure (BP) Medications	104027 (20.70)	4107 (61.31)	4681 (61.04)	996 (55.92)
Antithrombotic and Cholesterol	45525 (9.06)	4267 (63.70)	4958 (64.65)	1238 (69.51)
Antithrombotic and BP	42650 (8.48)	3582 (53.47)	4107 (53.55)	930 (52.22)
Antithrombotic, Cholesterol and BP	32374 (6.44)	3063 (45.72)	3532 (46.06)	794 (44.58)

Table 3 Medication History

BMJ Open

A Cohort Profile of the UK Biobank; diagnosis and characteristics of cerebrovascular disease

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-009161.R1
Article Type:	Cohort profile
Date Submitted by the Author:	02-Feb-2016
Complete List of Authors:	Hewitt, Jonathan; Cardiff University, Geriatric Medicine Walters, Matthew; University of Glasgow, Padmanabhan, Sandosh; University of Glasgow, Dawson, Jesse; University of Glasgow,
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Geriatric medicine, Epidemiology, Cardiovascular medicine, Neurology
Keywords:	Stroke < NEUROLOGY, EPIDEMIOLOGY, GERIATRIC MEDICINE

A Cohort Profile of the UK Biobank; diagnosis and characteristics of cerebrovascular disease.

Hewitt J[1], Walters M[2], Padmanabhan S[2] and Dawson J[2]

Jonathan Hewitt, MBBS, MSc, FRCP, PhD, Clinical Senior Lecturer and Corresponding Author

Mathew Walters, MD, MBChB, FRCP, Professor

Sandosh Padmanabhan, MD, PhD, FRCP, FAHA, Professor

Jesse Dawson, MD, MBChB (Hons), FRCP Clinical Reader

1. Department of Geriatric Medicine, Institute of Primary Care and Public Health, Cardiff University.

Address: Academic Building, University Hospital Llandough, Cardiff, CF64 2XX, Tel 02920 716982, Fax 02920 704244, Email hewittj2@cardiff.ac.uk, Corresponding author.

2. Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow.

Word Count 3249

A Cohort Profile of the UK Biobank; diagnosis and characteristics of cerebrovascular disease

Abstract

Purpose The UK Biobank is a large-scale biomedical resource, containing socio demographic and medical information, including data on a previous diagnosis of stroke or transient ischaemic attack (TIA). We described these participants and their medication usage.

Participants We identified participants who either self reported or were identified from a nurse led interview, having suffered a stroke or a transient ischemic attack (TIA) and compared them against participants without stroke ort TIA. We assessed their risk factor burden (sex, age, deprivation, waist to hip ratio (WHR), hypertension, smoking, alcohol intake, diabetes, physical exercise and oral contraception use (OCP)) and medication usage.

Findings to date We studied 502,650 people (54.41% women), 6669 (1.23%) participants self reported a stroke. The nurse led interview identified 7669 (1.53%) people with stroke and 1781 (0.35%) with TIA. Hypertension, smoking, higher WHR, lower alcohol consumption and diabetes were all more common in people with cerebrovascular disease ($p<0.0001$ for each). Women with cerebrovascular disease were less likely to have taken the OCP ($p=0.0002$). People with cerebrovascular disease did more exercise ($p=0.03$). Antithrombotic medication was taken by 81% of people with stroke (both self report and nurse led responders) and 89% with TIA. For self reported stroke, 63% were taking antithrombotic and cholesterol medications, 54% taking antithrombotic and anti-hypertensive medications and 46% taking all three. For the nurse led interview and TIA these figures were 65%, 54% and 46% and 70%, 53% and 45% respectively.

Future Plans

The UK Biobank provides a large, generalisable and contemporary data source in a young population. The characterisation of the UK Biobank cohort with cerebrovascular disease will form the basis for on going research using this data source.

Article Summary

Strengths and Weaknesses

A very large sample

Self Reported data

Key Words

Stroke, UK Biobank, Prevalence, Risk Factors

Introduction

The UK Biobank is a national health resource, intended to improve the prevention, diagnosis and treatment of illness [1]. It holds detailed demographic, social and medical information, along with physical measures, such as weight and blood pressure. Participants gave consent to have their future health records linked to their UK Biobank data, a process which is on going.

There is comparatively little epidemiological data on younger populations with stroke[2]. The UK Biobank obtained information regarding previous stroke or transient ischaemic attack and commonly associated medications (such as blood pressure and lipid lowering medications) in people aged 40-69 years old. It is an order of magnitude larger than previous self reported estimates of cerebrovascular disease. We aimed to describe the cohort of participants from the UK Biobank who self reported a diagnosis of stroke and transient ischaemic attack. We aimed to describe the medication data of the cohort.

Data Collection and Follow Up Between 2006 and 2010, the UK Biobank collected detailed data on 502,650 people. They used 22 assessment centres based in England, Scotland and Wales. Participants were recruited from NHS patient registers and contacted if they lived within a reasonable proximity to an assessment centre [3, 4].

Participants provided detailed demographic, socioeconomic and health related data via a touchscreen questionnaire, including medication history. Participants then underwent a range of physical assessment measures, included repeated blood pressure measurements, height and weight. Finally participants provided blood, urine and saliva samples which have been held in a purpose built central storage facility, in Stockport, UK, for future analysis. These depletable samples are due to be characterised by a comprehensive range of biomarkers by the end of 2015, but are currently not readily accessible.

One of the touchscreen questionnaire responses was “Has a doctor ever told you that you have any of the following conditions? (You can select more than one answer)” The possible touch screen responses were “Heart Attack, Angina, Stroke, High blood pressure, None of the above, Prefer not to answer”.

Following the questionnaire, a nurse led interview was performed to address past medical history in general. In order to prompt the nurses conducting the questionnaire and improve accuracy, all responses from the touch screen questionnaire were flagged to the study nurse. Therefore if a participant had selected “stroke” from the touchscreen question listed above then the nurse would have been aware of this response when discussing past medical history. The nurses also asked individuals directly about their past medical history. Therefore it is possible that someone may not have responded positively to the touchscreen questionnaire question regarding stroke but may have done so when asked later by the study nurse about past medical history. It is also during the nurse led questionnaire that participants would have been identified as having a past medical history of TIA, when questioned about their past medical history Further detailed information regarding conducting the nurse led questionnaire can be found at the UK Biobank website[5].

Participants were also asked to self-report via the touchscreen questionnaire their medication history. They were asked specifically if they regularly took “Medication for cholesterol” or “blood pressure medication” and could answer positively for both responses,

During the nurse led interview all medications where recorded as free text and subsequently grouped and coded by UK Biobank. We extracted data on reported antiplatelet or anticoagulant use. Any participant using any single or combination of these drugs was categorised as taking antithrombotic medication.

We also extracted demographic data concerning age, sex and socio-economic class, measured using the Townsend scale[6]. In order to select the most appropriate potential risk factors for stroke, we used data from the INTERSTROKE study [7] and the American Heart Associations' Stroke Councils Scientific Statement Oversight Committee guideline (AHA)[8]. Review of these publications identified seven modifiable risk factors that could accurately be reproduced from the UK Biobank. These risk factors represent the majority of the modifiable risk factors contributing to cerebrovascular disease.

From the INTERSTROKE study we identified hypertension, smoking, waist to hip ratio, regular physical activity, diabetes mellitus and alcohol intake as risk factors we could study. We could not include psychological factors, diet, cardiac causes and apolipoprotein levels as the data were not available in comparable format. From the AHA guideline we also identified the oral contraceptive pill (OCP). We used the same definitions as INTERSTROKE and the AHA guideline as far as possible; we defined a self-reported history of hypertension, smoking (current, ex or never), waist to hip ratio (divided into tertiles), moderate or strenuous activity for more than 4 hours per week and a self reported history of diabetes mellitus. For women, a history of ever having taken the oral contraceptive pill (yes or no).

Alcohol intake was based on the data recorded in the UK Biobank (alcohol daily or almost daily, alcohol on three or four days per week, alcohol once or twice a week, alcohol one to three times per month, special occasions only, never drinkers and former drinkers).

We describe the characteristics of the whole population and the population with stroke disease and TIA using descriptive statistics. We compared characteristics of patients who self reported stroke, those with nurse reported stroke and TIA against those who did not using t tests for continuous variables and chi-squared tests for dichotomous variables.The statistical analysis was conducted using STATA version 11 (Stata Corp).

Cerebrovascular Findings to Date

UK Biobank released data for use on 502650 participants. Their average age was 56.52 years, (standard deviation (SD) 8.09). There were 273468 women (54.41%) who were younger than the men (56.35 years (SD 8.00) compared to 56.74 years (SD 8.20), p<0.0001. The mean Townsend Index for all participants was -1.29 (SD 3.10).

More participants self-reported stroke during the nurse led interview (n=7669 (1.53%)) than on the touchscreen questionnaire (n=6699 (1.33%)). TIA was reported via the verbal questionnaire in 1781 (0.35%) of participants. Patients self-reporting stroke and TIA were older and more likely to be female (p<0.0001) in comparison to patients who did not. Higher levels of deprivation were seen in people self reporting cerebrovascular disease via both the touchscreen and nurse led questionnaire (p<0.0001) but this was not the case for reports of TIA (p=0.13). The full results are shown in table 1.

The cardiovascular risk factors in patients self-reporting stroke and TIA are shown in table 2 along with patients who reported no stroke or TIA.

Self-reported hypertension was recorded in 27.01% of the whole UK Biobank population. It was over 50% for each of the patients with a history of self reported stroke, stroke from the nurse led questionnaire and TIA. People with any cerebrovascular history were more likely to be former or current smokers. Diabetes was reported in 5.25% of the UK Biobank population but over 15% of stroke patients and 10% of patients with TIA. Each of these associations was significant (p<0.001, for each).

Self-reported alcohol intake was lower in people with cerebrovascular disease in comparison to the population without cerebrovascular disease. In people who drank less than weekly,

including never and former drinkers, there was a higher proportion of cerebrovascular disease recorded (p<0.001) compared to those who drank more frequently.

For women, there were 81.05% who had previously taken the oral contraceptive pill. For self-reported stroke and nurse led stroke responses, these figures were 73.83% and 74.22% respectively (p<0.001) and for TIA, this figure was 75.55% (p=0.0002).

There were 23823 (4.74%) people who reported doing at least 4 or more hours of moderate or strenuous physical exercise per week. Of the 303897 people with complete data for this response 302 (6.4%) of those who had a previous stroke identified via the touchscreen questionnaire responded positively to this question (versus 17458 (5.84%), p=0.1, for those without a previous stroke). For the nurse led self-report question these figures were 357 (6.47%) and 23463 (5.78%), p=0.03 and self-reported TIA they were 91 (6.45%) versus 17669 (5.84%), p=0.32.

In total, 76397 (15.20%) people who were identified as taking an antithrombotic medication, 104028 (20.70%) of people self-reported as receiving a blood pressure medication and 86907 (17.29%) of people self-reported receiving a cholesterol medication. For each of these responses participants who reported stroke or TIA disease all received more of these medications than the population as a whole, every response was significant (<0.0001). The full results are given in table 3.

Discussion

The objective of this study was to characterise the UK Biobank participants with a history of stroke and TIA and to gain insights into the treatment of patients at population level in the UK.

There are surprisingly few studies documenting the prevalence of stroke and TIA disease. For example, a 2012 review [2] identified only five studies of stroke prevalence in the Western world. They estimated that for an adult population, including older populations, the prevalence of stroke ranged from 0.15% in Italy to between 1.7% and 2.6% in the UK and the US. When these figures were assessed with respect to sex, there was a suggestion that prevalence was higher in men. Age is by far the largest non-modifiable risk factor for stroke making the findings in our younger cohort (1.74% for men and 0.99% for women) comparable.

Our findings noted a difference in reporting of prevalence between the self reported and nurse led figures for cerebrovascular disease . There were 970 more people identified with stroke disease following the nurse led questionnaire. Every person who responded positively on the touch screen questionnaire to having stroke disease will have been challenged by the nurse during the nurse led interview. All of these people were included in the nurse led responses as having a stroke, therefore this data has been at least partially verified by the nurse. The nurse also further identified people with cerebrovascular disease via questions on past medical history. We would advocate that the figure (7669) is the most accurate estimation of cerebrovascular disease within the UK Biobank.

Reduced socio-economic status was associated with both self and nurse reported stroke and TIA. Deprivation and stroke are associated with an increase in mortality across a range of countries[9] and also a Scottish meta-analysis linked deprivation with stroke incidence even after allowing for cardiovascular risk factors[10]. However, our data is an order of magnitude larger than the studies included in that review and adds a contemporary UK wide assessment of stroke and deprivation.

When selecting the risk factors to study from the UK Biobank we attempted to use the risk factors characterised in other studies. The advantages of this approach are that the reproducibility of our study is increased and our results can be compared directly to those generated from the high-income countries included in their studies. We were able to reproduce 6 of the risk factors used in INTERSTROKE study and the AHA guideline.

Hypertension was reported by 27% of the whole cohort but in over 50% of the participants with cerebrovascular disease. These figures are consistent with other epidemiological studies[11]. Our findings reflect both the causative effect of blood pressure and a heightened awareness of hypertension treatment in stroke survivors. Unsurprisingly our results show an increased number of current smokers and former smokers and a reduced rate of never smokers in those reporting cerebrovascular disease.

UK Biobank waist to hip ratio (WHR) data can be closely compared to the INTERSTROKE study data. This is because we were able to conduct our analysis using tertiles and also because while weight may have altered following a diagnosis of cerebrovascular disease it is likely to be one of the most challenging of modifiable risk factors to change following diagnosis, although data on post stroke weight change is sparse[12]. In the INTERSTROKE study, the incident rates were 23%, 33% and 43% for the increasing tertiles of WHR. These results are comparable to our results for both self-reported TIA (22%, 33% and 45% respectively) and stroke (18%, 30% and 51%). The high figure of 51% implies either an increased weight gain following stroke or a lower prevalence of obesity in the INTERSTROKE study, which included data from both high and low socioeconomic countries.

The self-reported history of diabetes mellitus was 5.25%. For a cohort of middle aged, predominantly Caucasian Europeans, this is what we would have anticipated[13] . Estimates put

diabetes prevalence between 10 and 25% in people with existing cerebrovascular disease depending on population and method used to diagnose diabetes[14] and our results (15% for stroke and 10% for TIA) compare closely to a self-reported diagnosis in a young population.

Our results suggest that people who undertook at least four hours of physical activity per week were more likely to have had a cerebrovascular event. Four hours is a relatively large amount and studies have consistently shown benefit in increased physical activity in reducing cerebrovascular disease[15]. Our results are likely to be due to reverse causation; either increased activity in cerebrovascular sufferers to improve their general health or the residual physical disability being reflected in normal activities of daily living being deemed at least moderately strenuous.

In people with cerebrovascular disease who drank 3 times or more per month, their frequency of drinking was below people without cerebrovascular disease. Further, in people who have never drunk or no longer drink there was more cerebrovascular disease, which may reflect either a reduced frequency of drinking following diagnosis or an increased prevalence of disease; the well described "J shaped curve" of alcohol related illness[16].

The oral contraceptive pill (OCP) is associated with ischaemic stroke, particularly at higher doses[17, 18]. We found that use of the oral contraceptive was lower for our cerebrovascular outcomes. Hence, it is possible that some of our participants with a past medical history of cerebrovascular disease were not prescribed the OCP, explaining the findings.

It is also possible that the result detected with cerebrovascular disease, physical activity, alcohol consumption and OCP use reflects bias associated with the cross sectional nature of the data and this need to be a stated limitation.

A high proportion of participants with self reported and nurse reported stroke (81%) and TIA (89%) were receiving an antithrombotic medication. The figure of 81% would seem appropriate for stroke, because up to 20% of people may have suffered intracerebral haemorrhage, thus restricting use and antithrombotic medication is not suitable for every individual. Our findings compare favourably to the Post Stroke Rehabilitation Outcome Project[19]. Their study in 2005 of 1161 people with both ischemic and haemorrhagic stroke, identified 62% of people as receiving antithrombotic medication.

While stroke disease does appear to have appropriate rates of antithrombotic prescription, TIA does appear to be under treated for antithrombotic medication. TIA is almost exclusively a disease results from infarction, embolic or otherwise, rather than haemorrhage. While antithrombotic will not be suitable for all of these people it is likely that 11% of untreated TIA disease represents suboptimal management.

We also demonstrated that large numbers of positive responders were taking cholesterol and blood pressure medications, including combinations of these medications. Not all of these medications will be suitable for each person with cerebrovascular disease, however, only about 45% of participants were receiving all three medications, suggesting these medications are under prescribed overall. Neither did we have information on whether treatment targets had been adequately achieved with these medications, something which is also likely to be suboptimal[20]. It is also likely that this young population of healthy volunteer participants are more likely to take and adhere to medication than less conscious populations.

It is important to highlight what we have not done in this study. Firstly, we did not consider all of the well recognised risk factors in as much detail as possible, for example, psychological disease. Smith and colleagues recently proposed criteria for probable major depression and probable bipolar disorder within the UK Biobank [21]. Currently these criteria

are not directly available from UK Biobank. As these and other criteria become established researchers will be able to test hypotheses directly between them and cerebrovascular disease in the UK Biobank. Secondly, we only studied our exposure and outcome variables in a comparatively simple form, when other possible combinations could have been considered. For example, we recorded self reported blood pressure but blood pressure was also recorded at the assessment visit. Neither did we perform detailed cross sectional analysis of each variable, such as smoking or alcohol and in particular the pharmacotherapeutic data. Both were conscious decisions. To have done so would have been outside the remit of our objectives, which were to provide an overview of the data. Each risk factor would in itself form a separate detailed analysis.

Thirdly we have not looked at many of the known but less well established risk factors. There were two reasons for not doing so. Often this was where the exact question did not exist within the UK Biobank, for example recreational drug use. The second reason was where clinical information was available it would be best studied in relation to future linked data. For example, will the people who self reported an asymptomatic carotid stenosis incur a different rate of incident cerebrovascular disease? Finally and perhaps most importantly this is self reported data. As such it is likely that it will under estimate the true prevalence of self reported and nurse reported stroke disease, as well as TIA.

Despite the limitations of the data, this sample of half a million people still represents the largest self reported estimate of cerebrovascular disease in a population of this or any age. It adds to the available data on young populations with self reported stroke and TIA disease which are less well described than older populations. The clinical relevance of the data is to highlight that many people were not receiving all of the secondary prevention medication that would have been expected.

Contributions, access fees, disclosures and data sharing statements.

J Hewitt conceived the project and performed all data analysis.

J Hewitt and J Dawson wrote the first draft.

J Hewitt, J Dawson, S Padmanabhan and M Walters developed the project, commented on results and contributed to all subsequent drafts.

The access fee for use of the UK Biobank was provided by the College of Medical Veterinary and Life Sciences, University of Glasgow.

None of the authors have disclosures and conflict of interest.

References

1. Collins R. What makes uk biobank special? *Lancet*. 2012;379:1173-1174
2. Zhang Y, Chapman AM, Plested M, Jackson D, Purroy F. The incidence, prevalence, and mortality of stroke in france, germany, italy, spain, the uk, and the us: A literature review. *Stroke research and treatment*. 2012;2012:436125
3. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, et al. Uk biobank: Current status and what it means for epidemiology. *Health Policy and Technology*. 2012;1:123-126
4. www.ukbiobank.ac.uk.
5. http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=100235.
6. Phillimore P, Beattie A, Townsend P. Widening inequality of health in northern england, 1981-91. *Bmj*. 1994;308:1125-1128
7. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the interstroke study): A case-control study. *Lancet*. 2010;376:112-123
8. Goldstein LB, Adams R, Alberts MJ, Appel LJ, Brass LM, Bushnell CD, et al. Primary prevention of ischemic stroke: A guideline from the american heart association/american stroke association stroke council: Cosponsored by the atherosclerotic peripheral vascular disease interdisciplinary working group; cardiovascular nursing council; clinical cardiology council; nutrition, physical activity, and metabolism council; and the quality of care and outcomes research interdisciplinary working group: The american academy of neurology affirms the value of this guideline. *Stroke; a journal of cerebral circulation*. 2006;37:1583-1633
9. Wu SH, Woo J, Zhang XH. Worldwide socioeconomic status and stroke mortality: An ecological study. *International journal for equity in health*. 2013;12:42
10. Kerr GD, Slavin H, Clark D, Coupar F, Langhorne P, Stott DJ. Do vascular risk factors explain the association between socioeconomic status and stroke incidence: A meta-analysis. *Cerebrovascular diseases*. 2011;31:57-63
11. Kannel WB. Blood pressure as a cardiovascular risk factor: Prevention and treatment. *JAMA : the journal of the American Medical Association*. 1996;275:1571-1576
12. Jonsson AC, Lindgren I, Norrving B, Lindgren A. Weight loss after stroke: A population-based study from the lund stroke register. *Stroke; a journal of cerebral circulation*. 2008;39:918-923
13. Group DS. Age- and sex-specific prevalences of diabetes and impaired glucose regulation in 13 european cohorts. *Diabetes care*. 2003;26:61-69
14. Hewitt J, Castilla Guerra L, Fernandez-Moreno M, Sierra C. Diabetes and stroke prevention: A review. *Stroke research and treatment*. 2012
15. Lee CD, Folsom AR, Blair SN. Physical activity and stroke risk: A meta-analysis. *Stroke; a journal of cerebral circulation*. 2003;34:2475-2481
16. Ferreira MP, Weems MK. Alcohol consumption by aging adults in the united states: Health benefits and detriments. *Journal of the American Dietetic Association*. 2008;108:1668-1676
17. Peragallo Urrutia R, Coeytaux RR, McBroom AJ, Gierisch JM, Havrilesky LJ, Moorman PG, et al. Risk of acute thromboembolic events with oral contraceptive use: A systematic review and meta-analysis. *Obstetrics and gynecology*. 2013;122:380-389
18. Zeitoun K, Carr BR. Is there an increased risk of stroke associated with oral contraceptives? *Drug safety : an international journal of medical toxicology and drug experience*. 1999;20:467-473
19. Zorowitz RD, Smout RJ, Gassaway JA, Horn SD. Antiplatelet and anticoagulant medication usage during stroke rehabilitation: The post-stroke rehabilitation outcomes project (psrop). *Topics in stroke rehabilitation*. 2005;12:11-19
20. McAlister FA, Majumdar SR, Padwal RS, Fradette M, Thompson A, Buck B, et al. Case management for blood pressure and lipid level control after minor stroke: Prevention

randomized controlled trial. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2014;186:577-584

21. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within uk biobank: Cross-sectional study of 172,751 participants. *PloS one*. 2013;8:e75362

		Sex		Age	Deprivation
Self Reported Stroke via Touchscreen Questionnaire					
		Men	Women	(Standard Deviation)	(Standard Deviation)
Yes	6699 (1.33)	3983 (1.74)	2716 (0.99)	60.8 (6.78)	-0.25 (3.54)
No	495951 (98.67)	225199 (98.26)	270458(99.01)	57.4 (6.78)	-1.25 (3.11)
Self Reported Stroke via Nurse Interview					
Yes	7669 (1.53)	4516 (1.97)	3153 (1.15)	60.89 (6.73)	-0.42 (3.48)
No	494051 (98.29)	224205 (97.83)	269852 (98.68)	56.4 (8.09)	-1.31 (3.08)
Missing (or not completed)	930 (0.19)	461 (0.2)	469 (0.17)		
Self Reported TIA via Nurse Interview					
Yes	1781 (0.35)	958 (0.42)	823 (0.3))	62.19 (5.87)	-1.31 (3.05)
No	499939 (99.46)	227763 (99.38)	272176 (99.53)	57.19 (7.9)	-1.22 (3.12)
Missing (or not completed)	930 (0.19	461 (0.2)	469 (0.17)		
Number (%)	502650 (100)	229182 (100)	273468 (100)		

Table 1. Self report for stroke and TIA with demographics

		Total UK Biobank	Stroke Diagnosed by touchscreen questionnaire	Stroke Diagnosed by nurse led questionnaire	Transient Ischemic Attack (nurse led questionnaire)
		N=502650 (100%)	N=6699	N=7669	N=1781
Hypertension (self reported)		135787 (27.06)	3770 (56.34)	4257 (55.51)	912 (51.21)
Smoking	Current	52989 (10.92)	1103 (16.48)	1223 (15.95)	191 (10.72)
	Former	173203 (34.52)	2851 (42.60)	3250 (42.38)	731 (41.04)
	Never	275565 (54.92)	2738 (40.91)	3196 (41.67)	859 (48.23)
Waist to Hip Ratio (tertiles)	Lowest	N/A	1181 (17.63)	1375 (17.93)	399 (22.4)
	Middle	N/A	1987 (29.66)	2308 (30.10)	574 (32.23)
	Highest	N/A	3470 (51.80)	3913 (51.02)	801 (44.97)
Self Report Diabetes	Yes	26408 (5.26)	1042 (15.57)	1151 (15.01)	178 (9.99)
Alcohol frequency	Daily or almost daily	101794 (20.29)	1217 (18.19)	1437 (18.74)	389 (21.84)
	3-4 times/week	115459 (23.01)	1100 (16.44)	1280 (16.69)	340 (19.09)
	1-2/week	129325 (25.77)	1560 (23.31)	1800 (23.47)	406 (22.8)
	upto 3/month	55871 (11.14)	713 (10.65)	821 (10.71)	200 (11.23)
	Special Occasions	58032 (11.57)	1010 (15.09)	1132 (14.76)	233 (13.08)
	Never	22551 (4.49)	476 (7.11)	529 (6.9)	108 (6.06)
	Former	18115 (3.61)	597 (8.92)	651 (8.49)	104 (5.84)
Oral Contraceptive Pill (N=272047)	Yes	220501 (81.05)	1989 (73.83)	2326 (74.22)	621 (75.55) (p=0.002)

Table 2, Descriptive Characteristics (p<0.001, unless stated)

	Total UK Biobank	Stroke Diagnosed by touchscreen Questionnaire	Stroke Diagnosed by nurse led Questionnaire	Transient Ischemic Attack (nurse led questionnaire)
	N=502650 (100%)	N=6699	N=7669	N=1781
Antithrombotic Medications	76397 (15.20)	5414 (80.82)	6243 (81.41)	1577 (88.55)
Cholesterol Medications	86907 (17.29)	4728 (70.58)	5466 (71.27)	1303 (73.16)
Blood Pressure (BP) Medications	104027 (20.70)	4107 (61.31)	4681 (61.04)	996 (55.92)
Antithrombotic and Cholesterol	45525 (9.06)	4267 (63.70)	4958 (64.65)	1238 (69.51)
Antithrombotic and BP	42650 (8.48)	3582 (53.47)	4107 (53.55)	930 (52.22)
Antithrombotic, Cholesterol and BP	32374 (6.44)	3063 (45.72)	3532 (46.06)	794 (44.58)

Table 3 Medication History