Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Hamburg University of Applied Sciences

Faculty of Life Sciences

Association between gender, Socioeconomic status, blood pressure, Body Mass Index, and stroke in Germany

An analysis based on the Robert Koch-Institute survey data (DEGS)

Master Thesis

Master Health Sciences

Submitted by

Cedric, Noune Tankou

[redacted]

Hamburg, 26th April 2021

First reviewer: Prof. Dr. Zita Schillmöller (HAW Hamburg)

Second reviewer: Prof. Dr. Sibylle Adam (HAW Hamburg)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BGS98	Federal Health Survey “Bundes-Gesundheitssurvey”
BMI	Body Mass Index
CI	Confidence interval
CAPI	Computer Assisted Personal Interview
CVDs	Cardiovascular diseases
DEGS	German Health Interview and Examination Survey for Adults
H_1	Alternative hypothesis
H_0	Null hypothesis
kg/m ²	Kiligramme per meter square
mm Hg	Millimeters of Mercury
NCDs	Non-communicable diseases
n.d.	No date
OR	Odd Ratio
p-value	Probability value
RKI	Robert Koch-Institute
SD	Standard deviation
SE	Standard error
SES	Socioeconomic Status
Sig.	Sigificance level
SPSS	Statistical Package for the Social Sciences
VIF	Variance Inflation Factor
WHO	World Health Organisation

Abstract

Background: The number of deaths due to NCD has increased from 2007 to 2017 by 22.7% (CI: 21.5-23.9). This represents an additional 7.61 million (CI: 7.20-8.01) deaths estimated in 2017 versus in 2007. Classified under NCD, a *group of heart and blood vessels disorders* called CVD occupies the first postion. CVDs are the leading cause of death globally and are responsible for arround 40% of deaths in Germany. Ranking CVDs in terms of mortality, stroke appears at position two behind coronary artery disease. 243 000 to 260 000 persons in Germany suffer from stroke each year according to RKI. Several risk factors could lead to stroke, of which the modifiables like blood pressure and the nonmodifiable like gender are registered.

Methods: Secondary data on the health status, health-related behavior, healthcare and living conditions of adults from the DEGS carried out by the RKI between 2008 and 2011 are used for a secondary analysis. The bivariate tests to assess the association between gender, blood pressure, BMI, SES, and stroke are mainly point-biserial correlation and chi-square, whereas a hierarchical multiple binary logistic regression analysis determines the predicting characters of these independent variables on the outcome stroke.

Results: Gender (X^2 (1) = 13.154, p< 0.001), high blood pressure (males: X^2 (1) = 44.714; p< 0.001; females: X^2 (1) = 52.019; p< 0.001), SES as a score (males: r = 0.054; p=0.008; females: r = 0.094; p<0.001) are positively significantly associated with stroke, however BMI is not significantly associated to stroke for males (r = -0.024; p=0.241), but negatively significantly associated to stroke for females (r = -0.070; p<0.001). Hypertension (males: b = -1.124; CI: 0.185-0.570; p < 0.001; females: b = -1.464; CI: 0.099-0.538; p=0.001), low SES compared to middle SES (males: b = -0.600; CI: 0.319-0.944; p=0.030; females: b = -0.675; CI:0.271-0.956; p=0.036) are significant predictors of stroke, whereas high SES (males: b = 0.324; CI:0.789 - 2.422; p=0.257 and females: b = 0.997; CI:0.809-9.082; p = 0.106) compared to middle SES and centered BMI (male: b = 0.032; CI:0.975-1.094; p = 0.273; females: b = -0.019; CI:0.927-1.039 p = 0.511) are not.

Discussion: The findings support the evidence that stroke depends on hypertension, score derived from SES for both males and females, BMI (for females) and gender, however, does not depends on BMI for males. Moreover, the predictors high blood pressure, low SES compared to middle SES had a significant influence on stroke for both genders, whereas BMI and high SES compared to middle SES had no influence on the outcome stroke.

Keywords: gender, SES, BMI, high blood pressure or hypertension, Stroke

1 Introduction

After conducting a systematic analysis for the global burden of disease study to assess the global, regional, and age-sex-specific mortality for 282 causes of death in 195 countries and territories, Roth et al. (2018) stated in their main findings that non-communicable diseases (NCDs) encompassed a greatest fraction of death and thus contributed to 73.4% (95% uncertainty interval [UI] 72.5–74.1) of the entire deaths in 2017, while the communicable diseases, maternal, neonatal, and nutritional (CMNN) on their own accounted for 18.6% deaths (17.9–19.6), and 8.0% injuries (7.7–8.2). Besides, the total numbers of deaths from NCD causes increased from 2007 to 2017 by 22.7% (CI: 21.5–23.9), representing an additional 7.61 million (CI: 7.20–8.01) deaths estimated in 2017 versus 2007 (p.1736). When NCDs as a whole is ranked, cardiovascular diseases (CVDs) do not only figure as the top one of NCDs, but as a leading cause of death globally and in Germany as well (Robert Koch-Institut, 2015a). CVDs are responsible for approximately 40% of deaths in Germany on one hand and on the other hand lead additionally to serious health consequences for both the individuals in terms of disability and for the society in terms of their chronic context which could lead to high medical costs or expensive therapies on the long term. As per defintion, a CVD is a group of heart and blood vessels disorders, of which the common ones are amongst other: coronary artery disease, heart attack and stroke (WHO, 2017). Between 2009 and 2019 as seen on Figure 1, the top ten rank causing death in Germany varies in an interchangeable manner (i.e. from colorectal cancer, through chronic kidney diseases till diabetes), however the top three most leading cause of death still evolve in a constant manner with stroke at the second position and a percentage change of 10.7 inbetween the decade (Institute of Health Metrics and Evaluation, 2019). There are several risk factors that could lead to stroke. These risks factors could be classified under two different group, thus, the modifiables (e.g high blood pressure, abnormal blood lipids, smoking, physical inactivity, obesity and diabetes) and the modifiables (e.g age, sex, hereditery) (Puthenpurakal & Crussell, 2017).

Since there are no such study neither analyzing the RKI health-monitoring data to assess the association between the independent variables (gender, Socioeconomic status (SES), blood pressure, Body Mass Index (BMI)), and the dependent variable/outcome (stroke), nor using a regression analysis model to assess the predicting character of these former variables on the outcome, the main aim of this study will therefore investigate this association and the predicting character of these variables. To be able to achieve this main aim, secondary data analysis based on the Robert Koch-Institute survey data (DEGS) recorded between 2008 and 2011 is carried out. Further analysis will compare to what extend the gender variable might be a susceptible risk factor of stroke on one side and on the other side will figure out how high blood pressure, BMI and each level of SES predisposes individual to stroke outcome.

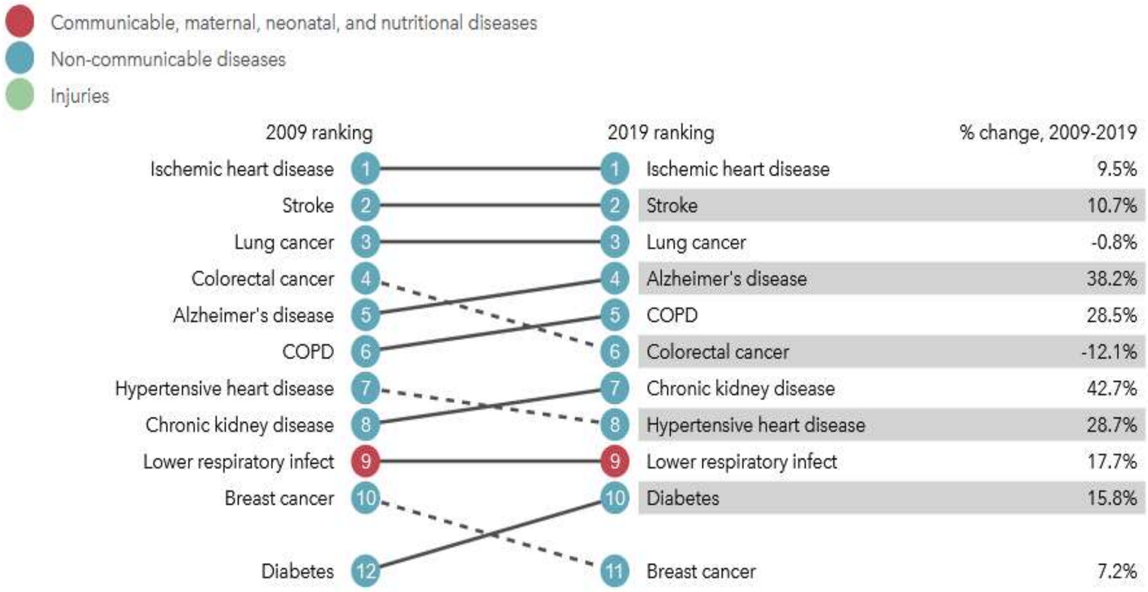


Figure 1: Most causes of death in Germany (Institute of Health Metrics and Evaluation, 2012)

To conveniently carry out this study, this essay will be written following a chronological order. To begin, the background and the public health relevance of the topics will be discussed, where amongst other the state of research and the research objectives, questions will be covered, then a focus will be laid on the methodology, in which the variables description and the data analysis process will be covered along with, followed by results from the different statistical analysis tests and after which, a discussion will be formulated, in which the study limitations and strength will be subjected to an adequate thinking, after which a final conclusion will formulated.

2 Background

2.1 Stroke Epidemiology

Stroke, also known as brain attack, is a vascular disease of the brain, commonly characterized by a sudden damage occurring on the brain tissues due to either a vascular obstruction or a cerebral hemorrhage. (Hankey, G., 2017). This abrupt damage characterizing the acute event occurs either posterior to an obstruction of blood vessels (ischemic stroke) by a blood clot or simply follows a rupture of the vessels (hemorrhagic stroke) (Hankey, G., 2016; Robert Koch-Institut, 2015a).

A transitory ischemic attack (TIA) in contrast, is a description of a neurological dysfunction that is fully reversible under the condition that it is handled within a period of 24 hours. Nonetheless, the term to be attributed to either form of the disease is based solely on clinical investigation findings. Ischemic strokes represent by far the largest subtype (around 80%), of which cardioembolic and microangiopathic or lacunary strokes are responsible for approximately 25% of this latter percentage (Erlanger Stroke Register, n.d., as cited in Mader & Schwenke, 2020), whereas primary brain hemorrhage and subarachnoid bleeding are significantly less common i.e. approximately 10-12% and 3% respectively. (Kolominsky-Rabas et al., 2015; Nimptsch & Mansky, 2012). Even thought the form hemorrhagic contitutes 20% of stroke, both forms remain nonetheless a global health problem responsible for the largest proportion of disability - adjusted life year loss and it is furthermore to be underlined that these forms account for two-thirds of deaths among all neurological disorders in the Burden of Disease Study in the year 2015 (Feigin et al., 2017; Plass et al., 2014).

Likewise globally, stroke is the second most common cause of death in Germany, the between 243 000 and 260 000 persons in Germany suffer from stroke each year (Robert Koch-Institut, 2015a). Though after conducting several studies, the statistics reported by different researchers vary from one source to another, they remain nonetheless alarming, then the prevalence of stroke indicates the percentage of stroke survivors within the general population and its high relevance for public health and health care planning as well. In the same vain, a study based on billing data from a national health insurance fund in Germany found a one year prevalence of 317 cases per 100 000 population for 2007 (Kohler et al., 2014), whereas in 2010 another study analyzed a German diagnosis-related groups billing data and the results revealed that, the hospital admission rate accounting for stroke was 365 cases per 100 000 population (Nimptsch & Mansky, 2012). Much more important to be considered as well is the fact that, those who have the chance to survive a stroke after hospital care, often have to deal with permanent damage issues. These sequels do not only reduce the individual's quality of life but also affect that of their respective family's member. One individual out of four will suffer from stroke as lifetime goes by, though either almost all ischemic or hemorrhagic strokes can be prevented by starting at each individual level to identify what could be the potential distinct risk factors (World Stroke Organization, 2019). Hence, to prevent stroke or to spare families from suffering from the damages which might result, one should act on the causes and particularly on the controllable factors.

Concerning stroke prevention, up to 90% of stroke cases could be prevented, if only several risk factors including hypertension, smoking and exercise were addressed. Preventive action on stroke could contribute by far to a massive scale reduction in stroke and enhance along with the global goals to reduce cardiovascular diseases and other significant causes of death occurring worldwide (World Stroke Organization, 2019). Some of the risk factors of stroke and state of research on these factors will be presented hereafter.

2.2 Literature Review and the Risk factors for Stroke

Risk factors could be responsible for an increase or a decrease in chances of getting stroke. Furthermore, they are biological, psychological, familial, communal, or cultural level characteristics that precede and are associated with a higher likelihood of negative outcomes (Substance Abuse and Mental Health Services Administration, n.d., p.1). Over three hundred risk factors are associated with coronary heart disease and stroke. The risk factors for CVD are significantly present in all populations. In the developed countries, the major five risk factors: tobacco use, alcohol use, high blood pressure, high cholesterol and obesity are responsible for at least one-third of all CVDs (WHO, n.d., p. 24). Risk factors could be classified under two categories namely: the non-modifiable and the modifiable factors. To influence positively or negatively his or her risk of contracting a disease, a person could actively control the modifiable factors, be it behaviors or exposures, whereas the non-modifiable are simply uncontrollable factors i.e., no action could be carried out, neither actively nor passively to get them changed in one way or the other. The most common risks factors of stroke differ from one source to another, however some of the modifiable risks are: hypertension, diabetes, smoking, physical inactivity and obesity, high cholesterol, alcohol consumption, cardiac causes (O´Donnell et al., 2010; American Heart Association, 2017), whereas age, gender and race/ethnicity belong to a non-exhaustive list of uncontrollable or non-modifiable (American Heart Association, 2017).

Regardless the factor sex, research shows that “Stroke affects both men and women, however, the incidence rates and outcomes differ between the 2 genders. Age-specific stroke rates are higher in men, but women experience more frequent stroke events because of their long-life expectancy, and high stroke incidence at older ages” (Reeve et al., 2008, as cited in Hiraga, 2017). Moreover, a systematic review leads to the results that: stroke is more common among men in the world, but the women are most severely ill; more precise result of the review showed that the mean age at which stroke occur in men was 68.6 years versus 72.9 years among women. Besides, when considering a large variations between age bands and between populations as well, the result moreover states that stroke incidence rate for males was 33% higher and stroke prevalence was 41% higher than that for the females (Appelros et al., 2009).

After conducting various other studies, different sources concluded that SES, high blood pressure, and BMI were associated with stroke, whereas other did not found this association. Along the same line, Busch et al. (2013) reported 2.9% to be the lifetime prevalence of stroke in Germany within the age group 40–79 years and stated that stroke prevalence increases continually amongst others: with increasing age and decreasing level of SES. Avan et al. (2019) on their side, used data extracted from the Global Burden of Diseases, Injuries, and Risk Factors Study with the main objective of analyzing them to figure out what were the trends in global and SES specific age-standardized stroke incidence, prevalence, mortality, and disability-adjusted life years lost between 1990 and 2017. The result revealed that the age-standardized rate of stroke incidence and mortality generally decreases in the respective SES level (derived from the subject’s income or socio-demographic index), although it should be further precise that this decline was much more exponential in the wealthiest societies (pp. 2, 26). Furthermore, a case control study included 347 ischemic stroke patients and 347 controls of the Guangzhou population to assess the effects of SES on risk of ischemic stroke. After adjusting for confounding factors, they came to the conclusion that high SES (derived from education, professional status and income) was positively related with risks of ischemic stroke; when components SES were detached and education expected closely, it was noticed that above 12 years of school attendance, the odds declined significantly to 2.18 (CI:1.25–3.82); odds of highest income was 2.83 (CI:1.25–6.39) compared to low income and odds of workers 1.87 (1.05–3.33) compared to the non-workers (Wang et al., 2019). In contrast to the preceding findings, another result from the association between these two variables SES and stroke was published, in which a negative association qualified by the researcher as inverse social gradient was indeed found i.e., an increasing prevalence of stroke was observed simultaneously with a decrease socioeconomic status level. When the gender variable was split into women and men, the lifetime prevalence of stroke was at its highest among people of low SES (much more clearly pronounced amongst women, 4.9% than men, 4.6%) and at its lowest amongst those of high SES (Busch et al., 2013). Both the income and occupational measurements of SES might also be linked to obesity, in the sense that, from a fortunate perspective of the SES, privileged individuals might tend to an overconsumption of food or tend to practice less or no physical activities, jogging because they can always afford their personal transport means, whereas the unfortunate would nourish themselves predominantly less adequately or, sometimes cover very long distances on foot, thus contributing to a reduction of their respective weights and consequently their respective BMI. Sedentary could also furthermore be seen amongst those subjects working during years in the office and do not need in one way or the other to perform any physical efforts, since all what they need (e.g., computer or telephone) in a daily or weekly basis fulfilled their various is just around them.

The association between BMI, high blood pressure, and stroke still knows controversy, although in the literature, a longitudinal study research, which included 26 607 elderly hypertensive subjects aged over 35 years in China, scientifically demonstrated that there was an association between BMI and stroke. They came to the results that, above the BMI threshold 30 kg/m², BMI was a highly significant and a fully independent risk factor for stroke both in men as in women (Heart rate=3.80, 95% CI: 2.47 to 5.86, p<0.001), compared with those having BMI between 18.5 and 24 kg/m² (Wang et al., 2013). Nonetheless, Dunbabin et al. (1990); O´Donnell et al. (2010), as cited in Furlan et al. (2018) after conducting a study on the relationship between arterial hypertension in the first 48 hours of stroke onset, reported that 80% of stroke cases are associated with high blood pressure. The source further states that high blood pressure could be responsible for hematoma enlargement production in hemorrhagic stroke on one side and in ischemic stroke, as well as could provoke cerebral edema particularly hemorrhagic transformation on the other side. In addition, hypertension could be classified top one of the modifiable risk factor for ischemic stroke. When the systolic blood pressure measurement of an individual crosses the threshold 160 mm Hg (Millimeters of Mercury) and/or his/her diastolic blood pressure measurement crosses 95 mm Hg, the individual is been considered as a person living with high blood pressure and this person might be four times more likely to be hurt from stroke, than those without underneath these mm Hg values (Sacco et al., 1997, as cited in Kumar, 2016). A review abstract of seven different studies which assigned each a relative risk of 1 for mild high blood pressure, determined the relative risk of stroke to be around the value 0.5, when the blood pressure was 136/84 mm Hg and 0.15 less when the blood pressure was 123/76 mm Hg (MacMahon et al. 1994, as cited in Kumar, 2016).

Further research on a woman health study including 39 053 subjects, used a Cox proportional hazards model to evaluate the association between BMI and stroke. After a mean follow-up of 10 years, a statistically significant trend for increased risk of total and ischemic stroke across seven BMI categories was found, thus, a total of 432 strokes (347 ischemic, 81 hemorrhagic, and 4 undefined) occurred. Since ischemic stroke constitutes by far the highest proportion of stroke, this risk was assessed in different multivariable models with total stroke (ischemic and hemorrhagic) as one variable and ischemic as another variable. The risk of total and ischemic stroke was significantly attenuated when antecedents of high blood pressure were controlled for in the various models. Furthermore, controlling for both diabetes antecedents and high cholesterol further decreased the heart rates values (and thus the risk) for total stroke and ischemic stroke as well i.e., 1.19 (95% CI 0.68 to 2.10) and 1.54 (95% CI 0.79 to 3.02) respectively and a figure was generated plotting their respective hazard ratio against their BMI. However, total stroke as a whole and ischemic stroke were differently associated with BMI: positive association for total stroke, however negative for hemorrhagic stroke. Additionally, a particularity revealed by the results of this study was that, women having a BMI of 20 kg/m^2 and less (Figure 2), had an increased risk of hemorrhagic stroke compared with women with a BMI ranging from 20.0 to 22.9 kg/m^2.(Kurth et al., 2005).

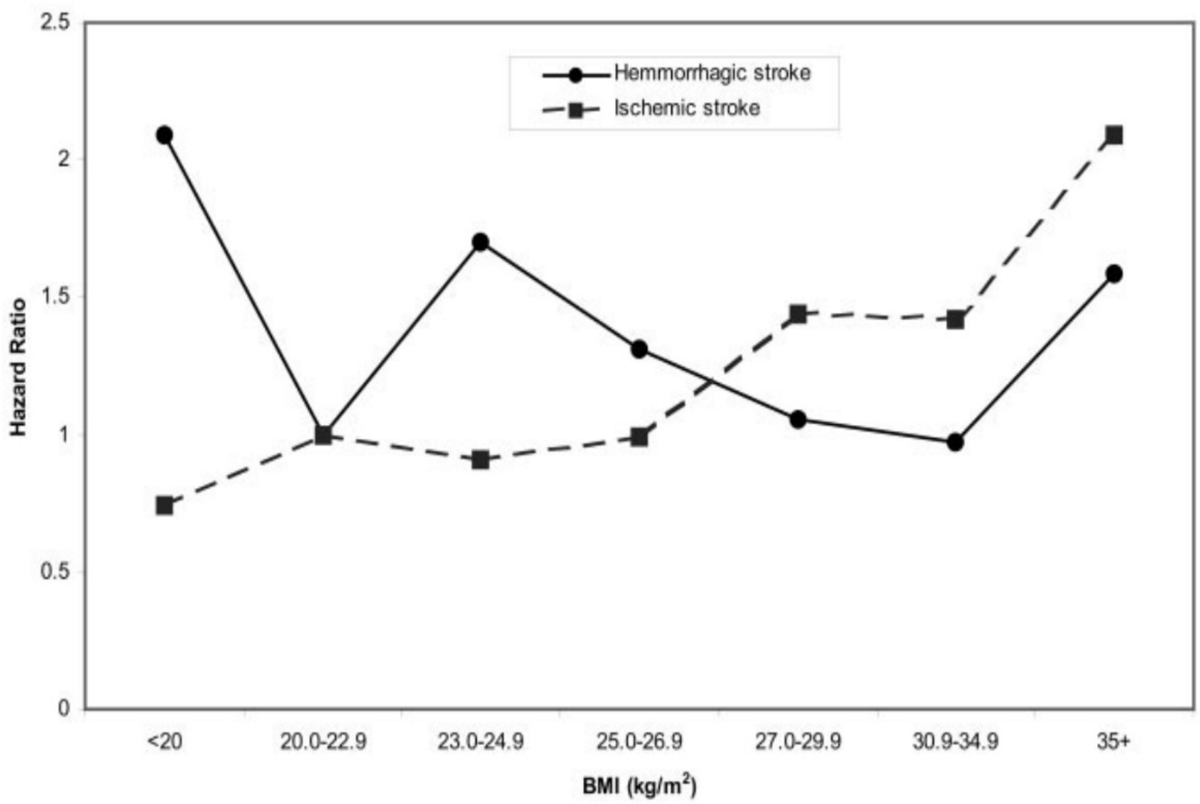


Figure 2: Multivariable adjusted heart rates for hemorrhagic stroke according to BMI categories, BMI of 20.0 to 22.9 kg/m2 serving as referent group (Kurth et al., 2005)

The literature review gives an insight overview on the state of research concerning some stroke risk factors and it is further noticeable that the study's results after several relationship assessments between the variables: BMI, SES, hypertension, gender, and stroke knows many contradictions. Since Robert Koch institute is the central institution of the Federal Government in the field of disease surveillance and prevention in Germany, the analysis of the data collected within the frame of its health monitoring program would permit to find out much more reliable representative results of the German population. The stroke risk factors BMI, SES, hypertension, and gender, (which are also explanatory variables in this study) will be discussed in detail here forth.

2.2.1 Body Mass Index

Body Mass Index is defined as a person's weight in kilograms divided by the square of the person's height in meters (kg/m2). It was formerly called the Quetelet index, and it represents a measurement for indicating nutritional status in adults (WHO, 2020). A specific Expert Consultation Group was constituted by the WHO in 1995 with the only goal to develop and assign uniform categories to BMI. The outcome this consultation yielded as a technical report in the same year in which four categories were attributed to Body Mass Index namely underweight, normal, overweight, and obese (WHO, 1995). A person would be considered as underweight if his index is below 18.5, as normal weight if his index range from 18.5 to 24.9, as overweight if his index is between 25 and 29.9, and as obese person, if his Index is greater than 30 (Table 1*).* Even though this is a good and current mean for an individual to calculate and figure out if height and weight are in a healthy relation, it must be mentioned that BMI is often criticized since in the arithmetic, it does not include an individual´s fat proportion and distribution. Besides, it is always not sure if self-reported height reflects the reality or if the reported weights are over or underestimated in such a way that the resulting BMI value ends up being mitigated (Lange & Finger, 2017).

According to Lange & Finger, 2017, though obesity is equally distributed over gender in Germany, men are far more overweight than women are i.e., respectively 43.3% in contrast to 28.8%, however approximately 35.9% of all adults are classified under the overweight category and 18.1% under the class I, II, III of obesity (Lange & Finger, 2017).

Irrespective under which nutritional classification an individual is classified, he/she will be exposed in one way or the other to different level of disease risks. These risks commodity vary from low to very severe as presented in Table 1 (WHO, 2000)

For several reasons, the classification of overweight and obesity is important and necessary as this permit meaningful comparisons of weight status within and between populations. The classification further serves as basis on which health interventions are evaluated on one side and on which individuals or groups with increased risks of morbidity and mortality concerning diseases are detected on the other side. Above all, based on this classification, priorities for intervention at individual and community levels could be set by researcher, health systems (WHO, 2000).

Table 1: Nutritional Status (WHO, 2020; WHO, 2000)

Nutritional Status	Body Mass Index	Risk of comorbidities
Underweight	Below 18.5	Low *Though increased risk of other clinical problems*
Normal weight	18.5 – 24.9	Average
Pre-Obesity (overweight)	25.0 – 29.9	Increased
Obesity Class I	30.0 – 34.9	Moderate
Obesity Class II	35.0 – 39.9	Severe
Obesity Class III	Above 40	Very severe

2.2.2 Socioeconomic Status

A few years ago, an analysis called Multiple Correspondence Analysis was used to create a wealth index based on 15 variables, mainly variables qualified as households' assets. These assets start with the type of housing, through the number of room in the household, the types of lightening these households were equipped with, emphasis on whether or not these households had features like kitchen, bathroom and toilet water supply and the assets finally end with assets called durables like for instance: television, satellite dish, phone and cell phone, stove, fridge, washing machine, and car (Engels et al., 2014). This index was much more concentrated on the belongings, properties, or assets aspect of SES, while the appellation SES speaks for itself in such a way that, in addition to the assets, should be further considered at least the interactions with the others or the social class of the subject concerned. This social aspect and several other concepts were measurements already considered while intellectualizing SES a century ago. The status was conceptualized by Taussig in 1920 as the occupational status of the father, while 14 years later, Cuff adopted a score card conceived by Sims in 1927 as a unit measure. This score card was all about interrogations on the household assets, on the parents' education and relevant information on the father's occupation (Taussig, 1920; Cuff, 1934; Sims, 1927, as cited in Broer et al., 2019, p.8).

Nowadays, the definition of SES does not seem to have change or evolve enough as it is been commonly used as the basics construct on which family background is being measured (Bofah & Hannula, 2017). Furthermore, it does not only represent the social standing or class of an individual or group, but "*It is often measured as a combination of education, income and occupation"*. Even though their inspections often reveal resources access inequities, privilege related issues or power and control problems (American Psychological Association, 2018), SES nevertheless "*constitutes a central analysis category of epidemiological research and health reporting. As part of the German Cardiovascular Disease Prevention Study 1984 – 1991, a multi-dimensional aggregated index was developed for the purpose of measuring SES*" (Lampert et al., 2013). As presented in Table 2, the RKI reported to have fundamentally revised the Index according to some critical assessments to ease health monitoring purpose.

The revised SES index appearing in the DEGS is based on three dimensions as well: education, occupation and income and they weigh equally in the resulting score. The dimension education of an individual is characteristic of an operationalization based on it academic and professional background, whereas the dimensions occupation and income in contrast are held to be characteristic of its household. Along this lane, continuous values are awarded from a minimum of 1 and a maximum of 7 points. The differences in point values should reflect differences regarding external criteria. That means, the metric scaling of the individual dimensions is a possibility (Lampert et al., 2013). The resulting categories and corresponding point values of the index are represented in Table 2. Even though it dates back to the latest 80s, Liberatos, Link & Kelsey state that there were several means to measure or apply a mathematical concept to the SES. These means depend especially a study´s context or the social class of it sample, on how specific measures are applicable to the population being studied, on the time at which a study is being assigned or even the indicators included in a study (Liberatos, Link, Kelsey, 1988, as cited in Broer et al., 2019, p. 8)

Table 2: Socioeconomic Status Index calculation basis (Lampert et al., 2013, p. 2)

Points	School and professional qualification	Professional status of respondents or of the head of household	Net equivalent income
1.0-1.9	– No school and no professional qualification (1a: 1.0) – Certificate of Primary Education ("Hauptschulabschluss") and no vocational qualification (1b: 1.7)	– Farmer: 10ha and more (1.0) – Farmer, no details provided (1.0) – Farmer: Under 10ha (1.1) – Unskilled workers (1.3) – Semi-skilled workers (1.8) – Workers, no details provided (1.9)	– ≤491€ (1.0) – 492–683€ (1.5)
2.0-2.9	– Certificate of Secondary Education ("Mittlerer Schulabschluss", "Realschulabschluss") or POS certificate and no vocational qualification (2b: 2.8)	– Foreman, group leader (2.0) – Skilled or specialist tradesmen (2.1) – Master, site foreman, overseer, (2.4) – Employees with executing responsibilities (2.4) – Others, no details provided (2.9) – Civil servants in Lower Service (2.9)	– 684–815€ (2.0) – 816–921€ (2.5)
3.0-3.9	– No school or primary certificate and training/apprenticeship/vocational school (1c: 3.0) – Certificate of Secondary Education, POS and training /	– Self-employed: no staff (3.5) – Employees doing qualified work (3.6) – Self-employed: 1–4 staff (3.6)	– 922–1082€ (3.0) – 1083–1188€ (3.5)

	appren-ticeship / vocational school (2a: 3.6) – Technical college qualification ("Fachhochschulreife"), University Entrance Qualification ("Abitur"), EOS and no vocational qualification (2c-gen: 3.7)	– Employees, no details given (3.7) – Self-employed in trading, business etc. (3.9)	
4.0-4.9	– Technical college qualification, University Entrance Qualification, EOS and training / apprenticeship / vocational school (2c-voc: 4.8)	– Self-employed or freelancer, no details given (4.0) – Civil servants in Intermediate Service (4.1) – Employees in a position of responsibility (4.2) – Self-employed: 5 or more employees (4.2) – Self-employed: PGH member (4.2) – Employees with extensive leadership responsibilities (4.7)	– 1189–1310€ (4.0) – 1311–1417€ (4.5)
5.0-5.9	– Category not taken	– Civil servants, no details given (5.0) – Civil servants in Higher Service (5.2) – Freelancers: no employees (5.8)	– 1418–1619€ (5.0) – 1620–1833€ (5.5
6.0-7.0	– Technical college qualification, University Entrance Qualification, EOS and Bachelor, Technical College Diploma (3a: 6.1) – Technical college qualification, University Entrance Qualification, EOS and Master	– Freelance academics (6.2) – Civil servants in Highest Service (6.4) – Freelancers: 1–4 employees (6.8) – Freelancers: 5 or more employees (7.0)	– 1834–2125€ (6.0) – 2126–2692€ (6.5) – ≥2693€ (7.0)

	/ Magister / Diploma, PhD (3b: 7.0)

POS: Polytechnic Secondary School ("Polytechnische Oberschule")
EOS: Extended Secondary School ("Erweiterte Oberschule")
PGH: Craftmen's Production Cooperatives ("Produktionsgenossenschaften des Handwerks")

2.2.3 Blood Pressure

Blood pressure is the ratio of the systolic pressure and diastolic pressure and stand for the force exerted by blood against the walls of the body's vessels in which it circulates. The systolic pressure represents the pressure in blood vessels when the heart contracts or when it beats, while diastolic is the pressure in the vessels when the heart rests between the beats. For hypertension to be diagnosed, the systolic blood pressure should be read two consecutive days and the values on both days should be greater or equal to 140 mmHg and/or the diastolic blood pressure should be greater or equal to 90 mmHg (WHO, 2019). These latter values are taken by the WHO to be the thresholds, in the sense that any number under these values are considered as absolutely normal (Table 3), even if Lewington et al (2002) bewail that an increase of both risks of stroke and coronary heart disease are detectable at even lower levels of 115mmHg for systolic and 75mmHg for diastolic pressure and that these risks increase steadily with each unit rise in blood pressure. Consequently, these thresholds levels could be seen as at the population level, just as alarming signals. The Blood Pressure Epidemiology Consortium (Konsortium zur Blutdruckepidemiologie) analyzed data from two national health surveys and five regional population-based studies which were conducted between 1997 and 2012, and came to the results that, blood pressure in Germany is still too high, though withing this period, it decreased for men as well as for women, particularly among individuals aged 55 to 74. Moreover, there were regional differences to be considered: blood pressure decreased far more in northeast region in contrast to the national average (Deutsches Zentrum für Herz-Kreislaufforschung, 2017). This latter source statements rejoins in one way or the other that of the authors Fehr et al. (2017), when they state that, the prevalence of high blood pressure in Germany is higher as compared to that of the European Union mean value: 20.2% for men and 21.7% for women. Since hypertension prevalence increases is with age, Fehr et al. go even further and link this high value to a demographic growth in Germany.

Up till the age of 44 years in women, high blood prevalence is below 10%, but in an exponential rising manner, this percentage attained approximately 55%, when the women reach 64 years and more. In men by contrast, the prevalence in younger ages is much more higher i.e. between 30 and 44 years the prevalence lies at 15%, whereas lies at 55% when the men are 60 years and above (Fehr et al., 2017).

"*Still too high"* reported the Consortium while referring to the burden of high blood pressure and the two words used by the WHO (2019) to qualify hypertension is "*silent killer*". Considering that hypertension often has no warning signs, the expression "*silent killer*" is being used her because most people live with hypertension and are neither aware of their health conditions regarding the diseased, nor are aware of the various risks or problems they might be exposed to. Hypertension favorizes besides heart attack, stroke, and kidney damage and that's why prevention should focus on health behaviors like the reduction of salt intake, the increase of physical activity amongst individuals and avoiding tobacco consumption (WHO, 2019)

Table 3: Classification of blood pressure levels (Tran & Giang, 2014)

Blood pressure classification	**Systolic blood Pressure in mm Hg**	**Diastolic blood pressure in mm Hg**
Optimal	Below 120	Below 80
Normal	120 – 129	80 – 84
High Normal	130 – 139	85 – 89
Hypertension		
Grade I	140 – 159	90 – 99
Grade II	160 – 179	100 – 109
Grade III	Above or equals 180	Above or equals 110
Isolated systolic hypertension	Above or equals 140	Below 90

2.2.4 Gender

Gender could be defined as "the array of socially constructed roles and relationships, personality traits, attitudes, behaviors, values, relative power and the influence that a society ascribes to the two sexes on a differential basis. Gender relational role and characteristic do not exist in isolation but are defined *in relation to one another and through the relationships between women and men, girls and boys*” (Health Canada, 2000), whereas sex is said to be nothing else than the biological differences existing between females and males. Still according to health Canada, the health sector focuses largely on the reproductive difference of the person i.e., its maternity aspect. Regarding the preceding definitions, it could be concluded that women and men are said to differ biologically, and the roles and responsibilities assigned to them by the society in which they live are neither alike as well. Besides, their various positions in their respective families and communities are different from one another tool. As a result of these differences, their conditions may be affected in one way or the other, the risks taken and exposure enhanced lead to a variety of outcomes, their efforts to improve their health have to increase, and the health system responses to their needs are affected (World Health Organization, 2020). Gender distribution according to age groups in Germany is illustrated in Figure 3.

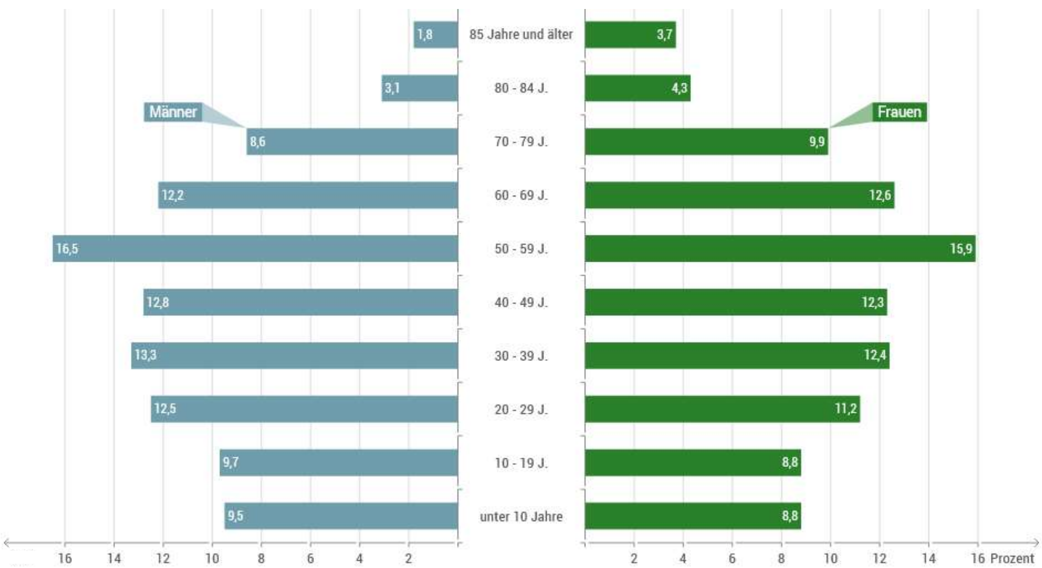


Figure 3: Gender distribution in Germany according to age groups (Bundeszentrale für politische Bildung, n.d.)

The research and findings published on stroke so far relate to the investigational aspect of the disease carried out on scientific basis, however the clinical aspect should not be ignored as it put forward the clinical picture (which could be sometime scaring) in terms of signs and symptoms that are provoked by an illness and/or individuals' reactions to the disease. Besides this, there is also an after-event phase, pronominally, for those, who, due to an intermittent care, treatment etc. survive the event. A phase sometimes referred to as rehabilitation, where the main objective is to overcome the sequels and regain some sort of patient autonomy as referred hereafter.

2.3 Clinical Features

According to Delamaire & Lafortune, 2010, as cited in Robert Koch-Institut, 2015a, a total of 58,556 people (35,389 women and 23,167 men) died from cerebrovascular disease (ICD-10: I60–I69) in 2013. This corresponds to 6.6% of all deaths (7.4% for women, 5.4% for men). Cerebrovascular disease is the most important expressive or clinical form of stroke manifestation.

The acute complaints and symptoms of stroke can be very different and depend on the type and the extent of damage, which would have occurred, on the affected brain region, and finally will depend on whether an individual is a male of a female. However, typical non-exhaustive symptoms are, sudden numbness or weakness in the face, arm, or leg, especially on one side of the body; rapid confusion, trouble speaking, or difficulty understanding speech; abrupt trouble seeing in one or both eyes; unexpected trouble walking, dizziness, loss of balance, or lack of coordination and hasty severe headache appearing suddenly or un-suddenly especially with no known antecedents causes. Once one or more of these signs occur, every minute counts, thus acting quickly and actively for an eventual imminent care is the only key to an eventual recovery (Centers for disease control and prevention, 2020).

2.4 Effects and Life Conditions after Stroke Event

Epidemiology is by definition, the scientific, systematic, and data-driven study of the distribution (frequency, pattern) and the determinants (causes, risk factors) of health-related states and events in specified populations (Centers for disease control and prevention, 2016). That means, the epidemiology of stroke often illustrates or captures the incidence, the prevalence or even the morbidity of the disease, however regarding this definition, the aspect of life after the disease events it not quite clearly underlined, although after stroke event, the repercussions or sequels could be as much as alarming, severe or even worst as the disease itself.

That could sometimes explain why in some cases, to be able to definitely relieve or stop pain, euthanasia is being practiced on patients suffering from uncurable diseases or perhaps in case their conditions turn out to be hopeless on a medical point of view. For more facts on the disabilities, stroke has imposed 132.1 million disability-adjusted life years (DALYs) lost globally in 2017 (this represent 34% more as compared to the early 90s), 42% of which was related to ischemic strokes, more precisely, 6.8 million DALYs in low income countries, 47.1 million DALYs in lower middle income countries, 63.1 million DALYs in upper middle income countries, and 14.2 million DALYs in high income countries (Avan et al., 2019).

After stroke event, patient usually suffers from the disorders, of which the most common ones apart from hemiplegia are, speech, swallowing, vision, and balance disorders, as well as consciousness and perception disorders. According to the National institute of neurological disorders and stroke (2020), these disorders could be classified under five types of disabilities. The first disability is: paralysis or problems affecting movement control (motor control) which expresses itself by hemiplegia or unilateral paralysis. The paralysis may affect the arm, hand, leg, or the entire half of the body in such a way that patients are often not any longer (or just partially) able to perform everyday basic activities such as washing, going to the toilet, or climbing stairs. The second disability concerns sensory disturbances where the individual loses ability to feel objects, pain, touch; and the third disability deals with difficulties in using or understanding language (aphasia). Aphasia occurs when the language and the comprehension faculty are attained. Moreover, if the language center in the brain is damaged, both speaking and understanding or even reading and writing capacities could be affected. The next disability is a problem linked with thought and memory, which could be manifested by an anosognosia, defined as "*an inability to acknowledge the reality of the physical impairments resulting from a stroke*" (National institute of neurological disorders and stroke, 2020). The final disability is the emotional disturbance, which is characterized by depression and other psychological consequences. This happen when stroke injures the brain and thus, may have a direct effect on the patient's emotions. Besides, restlessness, impulsiveness, and aggressiveness may also occur. In the same vein, those affected may not always be able to recognize or understand the consequences of their own situation and that is why they often experience anxiety, despondency, exhaustion, and depressive moods while coping with the disease (Stiftung Deutsche Schlaganfall-Hilfe, n.d.). As a result, living with stroke resumes itself by trying in all possible means of rehabilitation to regain at least the far minimum possibility of independency and autonomy.

On one hand, the narration in the background states the magnitude particularly the public health relevance of the disease stroke all over the world and in Germany in particular, whereas, on the other hand, the actual research status on stroke is highlighted by the literature review (*section:* 2.2). Since some research results regardless their publication dates, are basically controverting or contradicting other results, it is therefore a signal that the whole thematic has not yet been covered and it is therefore imperative that new information should be generated by other studies contextualized differently, thus, the necessity of the conduct of this study.

In view of the above stroke epidemiology and literature on stroke risk factors, it follows that, to conduct this study, an approach in which research aims, questions, objectives and hypothesis prevail and are accordingly derived and contextualized in advance should be envisaged as presented henceforth.

2.5 Research Objectives and Questions

Via personal interviews, physical examinations and analyses of blood and urine samples, the RKI within the frame of DEGS survey collected detailed and valid information for public health and health policy purposes (Robert Koch-Institut, 2015b). As a result, this data collection approach leads to more valid measurements and permits a better estimate of disease prevalence as well. This data collected over Germany is the most suitable for a secondary data analysis of with this current study and hence, suitable for assessing the relationship between the independent variables gender, SES, blood pressure, BMI, and the outcome stroke and as well suitable for an assessment of these variables predicting character. To achieve this, the following research aims, objectives and research question are formulated.

2.5.1 Research Aims and Objectives

To effectively conduct this study, the research aims, and objectives formulations are illustrated in Figure 4 below.

Research aims:

- Assessment of the association between gender, socioeconomic status, blood pressure, Body Mass Index, and stroke in the German context.
- Assessment of the strength of association between gender, socioeconomic status, blood pressure, Body Mass Index, could be associated to stroke.
- Assessment of the prediction of stroke by the variables: gender, socioeconomic status, blood pressure, Body Mass Index.

Research objective 1

To explore how each level of SES predisposes individual to stroke outcome.

Research objective 2

To generate new information for the effective improvement of existing treatments and adapt care policy approach.

Research objective 3

To raise awareness since the knowledge and tools available could be a limited factor or an inadequate feature to tackle a disease.

Research objective 4

To derive new information that could clarify the contradicting results of other research and propose recommendation to improve effectively existing public health approach and adapt care policy.

Figure 4: Research aims and objectives formulations

2.5.2 Research Questions and Hypothesis

From the prior aims and objectives, the following research questions, null (H_0) and alternative (H_1) hypothesis formulations of this study have been derived and presented in Figure 5 below.

Research question 1

Is there a gender, SES-Score, blood pressure, and BMI difference in stroke?

Hypothesis

H_0: There is no gender, SES-Score, blood pressure, and BMI difference in stroke.

H_1: There is a gender, SES-Score, blood pressure, and BMI difference in stroke.

Research question 2

Is there a correlation between gender, SES-Score, blood pressure, BMI, and stroke?

Hypothesis

H_0: There is no correlation between gender, SES-Score, blood pressure, BMI, and stroke.

H_1: There is no correlation between gender, SES-Score, blood pressure, BMI, and stroke.

Research question 3

Do gender, SES, blood pressure, and BMI predict stroke?

Hypothesis

H_0: Gender, SES, blood pressure, and BMI do not predict stroke.

H_1: Gender, SES, blood pressure, and BMI predict stroke.

Figure 5: Research questions and hypothesis formulations

For this research to be scientifically conducted and for the research aims and objectives, through the various research questions and hypothesis formulations to be attained, it is imperative that a specific procedure should serve as the basis on which the research is carried out. Hereafter, the research methodology, including amongst others the data source, sampling and data collection process is presented.

3 Methodology

3.1 Generalities and Data Source

To answer the research questions, a secondary data analysis based on routine data from the German Health Interview and Examination Survey for Adults "Die Studie zur Gesundheit Erwachsener in Deutschland" (DEGS), published in 2015 is performed. The DEGS survey is part of the health monitoring (Appendix VIII) program at the Robert Koch-Institute (RKI) in Germany, where representative data on the health status, health-related behavior, healthcare and living conditions of adults were repeatedly collected using appropriate methods. The first data collection and thus, wave 1 (DEGS1), was conducted from 2008 to 2011 and was designed as an interview, examination and test (Robert Koch-Institut, 2015b). The DEGS1 mixed design, permits both cross-sectional and longitudinal analyses and included the resident population of Germany aged 18–79 years. The target population of the DEGS1 were Germans living the Federal Republic of Germany at the period at which data were collection and as well included, foreigner within the same age group which could prove by an administration attest that their main residence was in Germany. The total sample size included in addition to the new recruited aged from 18 to 79 years, the former participants aged between 28 and 91 years old of a 1998 study named Federal Health Survey "Bundes-Gesundheitssurvey" (BGS98). A total of 8,152 women and men aged between 18 and 91 participated in DEGS1 study. The response rate within the DEGS1 study was 42% whereas that within the BGS98 study was higher (62%). The data from this survey was said to be adequate for this research since they provide information on the most widespread diseases, health risk factors and healthcare problems in Germany (Robert Koch Institut, 2009, 2013).

3.2 Sampling Design and Description

Two stages characterize the sampling design of the DEGS1 survey in Germany. Within the first stage, a total of 180 sample points, roughly and evenly distributed across the whole country were chosen (Figure 6), of which 120 sites from the former BGS98 study and an addition of 60 sites from the DEGS1 itself (Kamtsiuris et al., 2013). The DEGS1 study was conducted between 2008 and 2011 and the new recruited had an age ranging from 18 to 79 years (Robert Koch-Institut, 2015b, p.1), whereas the BGS98 study was carried out from October 1997 to March 1999, in which the participants were more older with an age range between 28 and 91 years (Thefeld et al., 1999, p.201).

This increase of the sites would have been purposeful in the sense that, it would have increased the statistical power on one hand and on the other hand would have adjust the sample to that of the population structure as it was 10 years ago (i.e., back in the 1990), so far, the significant demographic shift, which occurred in some part of Ost-Germany is considered.

In contrast to the BGS98 study, where the subjects were invited to take part in the survey, the selection of the resident’s, participant´s addresses for the DEGS1 survey was performed by a mathematical random procedure in their registration file provided by the authorities and the resulting data was adjusted by age (Koch-Institut, 2009, pp. 88-81).

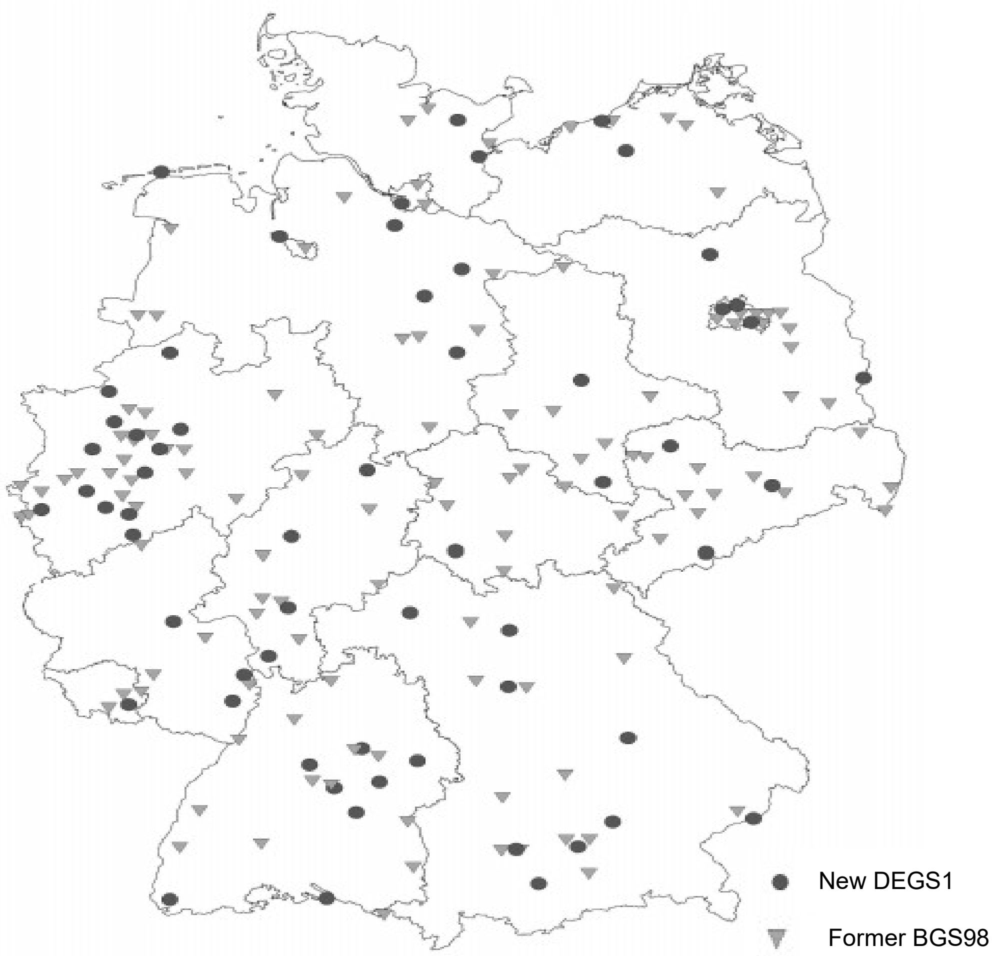


Figure 6: Former BGS98 and New DEGS1 sampling location over Germany (Koch-Institut, 2009, p.80)

3.3 Data Collection Process

Apart from questions on the sections: physical examinations, medical interview, the program emphasized on other tests like for instance laboratory testing of blood and urine. Besides, the medical examination included weight and height, blood pressure and pulse, thyroid gland size and aspects of what could be qualified as physical activity. Furthermore, the results of blood and urine analyses were used to determine health risks, since either physical examination nor self-reported data on a questionnaire form could not derive information like nutrient deficiency, allergic sensitization, or eventual cardiovascular disease risk. A report with the measured laboratory parameters was issued after all examination sessions and the participants were recommended to consult a physician in case of abnormalities in the findings (Robert Koch Institut, 2013, pp. 1-2).

3.3.1 Survey Instruments

The instruments used in the survey to collect the diverse data could be presented under three main categories: questionnaire, medical interview (in a Computer Assisted Personal Interview (CAPI) form and measuring form. Several variants of these instruments are presented in Table 4 below. More detailly, the self-filling questionnaire contents were based on essential health-relevant questions, such as, their respective quality of life, their respective opinion about the different services offered by the health system, and concerns about their working and environmental conditions. In contrast, the standardized medical interview included questions about diseases considered to be infectious and besides, eventual subject´s vaccinations. The physical examinations whereas focused on basics, such as the anthropometric measures (size, weight, hip circumference, waist circumference), blood pressure and pulse measurement, sonography of the thyroid gland. There was also laboratory test, where the information or results from the questionnaires, the medical interview, the drug interview, the physical measurements, and tests were being supplemented and rounded off by the laboratory blood, serum, urine, and saliva analyses. Finally, the standardized drug interview was conducted by the medical staff of the RKI with the help of a tool call “Instruments zur Datenbank gestützten Online-Erfassung von Medikamenten (IDOM)” or online supported database recording of drugs (Robert Koch Institut, 2009, pp. 69-78).

Table 4: DEGS1 survey instruments (Robert Koch-Institut, 2015b, p. 1; Robert Koch Institut, 2009, pp. 69-78)

Instruments	Variants
Questionnaire *(Filled by the participants)*	• 18-64 years • 65+ years • short questionnaire • Nutrition/Food Frequency Questionnaire
Medical interview (CAPI) *(Standardized)*	• CAPI • Short CAPI • Telephone CAPI
Measuring form	• 18-64 years • 65+ years

3.4 Independent/Predictors Variables

This section is reserved for an overview description of the different variables included in the analysis of this study as collected by the RKI within the frame of the study DEGS1.

3.4.1 Gender

Gender was collected and coded as 1 for men and coded as 2 for women. For analysis in this study, the variable gender is used as it was collected and recorded in the DEGS data set.

3.4.2 Socioeconomic Status

Conceptually, SES is used to describe the individual position in an unequal society. Within the framework of health research in Germany, multiple suggestions have been made to operationalize SES. (Lampert & Kroll, 2009, as cited in Robert Koch-Institut, 2015b, p.12). According to the SES Index calculation as presented in Table 2, information on the variable SES were calculated based mainly on the three equally weighed dimensions: education, occupation, and income. Additionally, information on education and training was classified primarily according to the International Standard Classification of Education (ISCED) classification system

(Schroedter et al. 2006, as cited Robert Koch-Institut, 2015b, p.12) and secondarily according to the Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) educational classification system (Lechert et al. 2006, as cited Robert Koch-Institut, 2015b, p.13). The resulting continuous variable, thus SES total score, obtained from the addition of the subs-scores: occupation, education, income (*section 2.2.2)* is used in this study for subsequent descriptive statistics and bivariate analysis as well. Moreover, within the framework of the DEGS1 research and more precisely following the whole SES calculation process, the resulting total scores were finally computed as an ordinal variable in the order: low, middle, and lastly high SES. Since, ordinal scaled variables have neither appropriate defined scales nor fixed intervals, they may not be fitting in a regression model equation as predictor and hence, to adapt their categories to these regression models, they must undergo in advance a transformed within a coding process described as dummy coding (Manfred & Paula, 2015, p. 3). In other words, dummy variables get around the assumption that the predictors level measurement in a model should be on a scale measurement. In the context of this study, this ordinal categorical variable is dummy coded and middle SES set as the reference group.

3.4.3 Blood Pressure

Systolic and diastolic pressure depend interchangeably on one another. To derive the average blood pressure of the subjects, the formula: $p * SYS + (1 - p) * DIA$ was applied. Where p is the mean value of three different consecutive measurements of the pressures for all DEGS participant, SYS is the systolic blood pressure and DIA is the diastolic blood pressure. For all participants in the DEGS1 study, the mean alpha value for each of the three repeated measurements (systole, diastole, and blood pressure) were calculated, and the results means yielded were 0.396, 0.3913 and 0.3901. In case a participant has a measurement value with a deviation of more than 0.25 from the alpha mean value within the control or review phase, this was checked up to figure out where the error might have come from and where possible, corrections would have been made accordingly. Once all deviations have been reviewed and corrections made, but some data are still despite all the preceding process implausible, then the respective values from systole, diastole or arterial pressure would have been set to Missing (Robert Koch-Institut, 2015b, p. 6).

3.4.4 Body Mass Index

BMI was calculated respective to the weights (in kg) and heights (in m^2) of the participants. Whenever for one reason or the other, a participant could not be examined according to the standardized preestablished protocol, the reasons would have been noted in comments for further eventual investigation or subsequent considerations during data manipulation. For instance, under certain circumstances, corrections would have been made on a subject weight and/or height, if it happens that a subject weighed his/herself with cloths or when he/she measured his/her height while having his shoes on. In these preceding two cases, 1 kg would have been subtracted from the weight and 3 cm automatically deducted from the height. As for too heavy subjects i.e., weighing over 200 kg, their weight values could not be measured, but were nevertheless included in the study, in case they knew and could communicate their respective weight to the investigators (Robert Koch-Institut, 2015b, p. 6). The BMI values end up being recorded as a metric variable and are also analyzed in this study as such, however, are computed into a nutritional status, more precisely on an ordinal categorical level for descriptive analysis purpose (Table 1).

3.5 Outcome variable

The only outcome variable in this study is stroke. Stroke was recorded as a categorical dichotomous variable. During a Computer Assisted Personal Interview, the subjects were asked directly independently of any form of medical consultation, if they were already diagnosed with stroke in the past or not. In case they were, the investigators dived in the antecedents of the participant to acquire further information that might be helpful in course off the DEGS survey. At this stage, it was therefore all about a doctor's diagnosis of stroke as reported by the subject (Robert Koch-Institut, 2015b, pp. 6-7).

3.6 Statistical Analysis

For analysis and meaningful results interpretation purposes, some data needed for this study have been either recoded or computed into new variables. Since all data are aggregated, but not directly recorded on an individual or study participants, neither patient consent nor ethics committee approval had to be requested before conducting this research. All statistical analyses are carried out using IBM Statistical Package for the Social Sciences (SPSS) Version 25. Moreover, the level of significance (sig.) is taken to be 95% and an error probability alpha ($\propto$) of 0.05.

To exclude any possible distorting effect on results, and for further comparison purposes, all statistical analyses except the one investigation the variable gender, are split by gender.

3.6.1 Descriptive Statistics

The measurements used for the description of variables are accordingly: variation, central tendency, and distribution. Moreover, the sample of this study is described by gender, BMI, SES score, SES. While BMI is categorized into four groups i.e., underweight, normal weight, overweight other pre-obesity and obesity, SES is categorized into an ordinary variable with the levels low, middle, and high. Both stroke and high blood pressure are described by their prevalence observed in the sample, whereas the continuous variables in contrast are described by their mean, standard error (SE), standard deviation, variance and respective minimum and maximum. Whether the continuous variables (BMI and SES score) fulfilled the assumption of normal distributed or not, this is tested by the Kolmogorov-Smirnov test (Field, 2018, p. 249). Besides, Cross Tabulation is used to explore the difference in stroke between two categorical binary variables (gender and blood pressure) (Joel C., 2017). Since the two continuous variables could not show a normal distribution, U Mann Whitney T test aided to test and confront the null hypothesis (H_0) that there is no SES score, and BMI difference in stroke with it corresponding alternative hypothesis (Milenović, 2011). The large sample of this study made it possible for all analysis (inclusive the descriptive statistics) to be executed per listwise. Certainly due to this enormous sample size and because independent sample t-test is a robust test, this latter test is used to assess the SES score and BMI difference in stroke (Statistics Laerd, 2018), though a non-parametric was the ideal with regard to the unnormal distributed data (Milenović, 2011).

3.6.2 Bivariate Analysis

Since the DEGS1 dataset contains too much missing values and enough valid subjects as well to fit the analysis of this study, there was no need for a lager sample size per pairwise exclusion to be guarantee. Therefore, cases are selected per listwise exclusion throughout all bivariate analyses. The tests are furthermore two-tailed so that the effects observations could be done in both directions. Kolmogorov-Smirnov test for the variables SES score and BMI yielded a highly significant result ($p < 0.001$), signifying that these variables are not normal distributed. The normality plot does not influence that much the tests of correlations within the frame of this study, then the nature of the variables does not fulfill the assumptions to run a conventional Pearson correlation test, but rather, fulfill the assumptions of another test, namely, a non-parametric. In the same vein, Field (2013) states that, to assess the correlation between the continuous and dichotomous variables the point-biserial correlation is chosen which is done with the Pearson's correlation coefficient (r) (pp. 279-280), since SPSS does not have a special procedure to be able to run the *Point-Biserial Correlation Analysis.* Further dichotomous variables are analyzed with the φ- coefficient which basically relies on the chi-square (Field, 2013, pp. 725, 740). Brosius (2013) affirms that the coefficient yielded by a chi-square test result, would be typically undistinguishable with the coefficient yielded by a Pearson coefficient output (p. 433). The guidance of this latter source is used for the interpretation of the correlation coefficients subsequently to the analysis. This guidance to coefficient interpretation is illustrated in the Table 5 below.

Table 5: Brosius correlation coefficients Guidance (Brosius, 2013, p. 523)

Correlation coefficients	Strength of correlation
0	None
Above 0 till 0.2	Very weak
Above 0.2 till 0.4	Weak
Above 0.4 till 0.6	Moderate
Above 0.6 till 0.8	Strong
Above 0.8 till less than 1	Very strong

3.6.3 Hierarchical Multiple Binary Logistic Regression

Conferring field (2018) the variance of an equation could be wrongly interpreted, if pairwise exclusion is ticked off during a test execution, as this exclusion modifies the sample and the value of the variance of the outcome variable along with each predictor included in the model (p. 408). As a result, cases are excluded per listwise throughout this analysis to prevent incongruities, particularly absurdities.

To find out whether gender, SES, SES score, blood pressure, and BMI are stroke predictors, a hierarchical multiple binary logistic regression analysis is carried out. This is the appropriate model, on one hand, because the dependent variable is binary categorical (stoke) and on the other hand because this analysis is powerful enough and capable of controlling potential confounding factors. Despite the robustness of this multiple binary logistic regression analysis, potential confounding factors will be nevertheless added to the analysis to minimize any possibility of ending with a biased result. Hence, this test is run to predominantly determine the change of the variability explained as one move from one model to another by adding each time a variable, including the confounders themselves. A confounding variable is independently related to two variables of interest that falsely obscures or accentuates the relation between the two variables in question (Meinert & Tonascia, 1986, as cited in Maartje et al., 2010). A confounder could be indeed similar to a mediator i.e. a variable that accounts for the relation between a predictor and an outcome, however a confounder is not intermediate in a causal casual sequence like a mediator is (Baron & Kenny, 1986, p.1176). As far as the suppressors or confounders are concerned, Li et al. (2019) conducted a cross sectional study on 31,464 subjects divided into low risk, moderate risk, and high risk with a specified objective to investigate "*the risk of stroke and its associated risk factors in a health examination population*". The results revealed that, the percentage of subjects exposed to stroke increases in a highly significantly manner (p <0.001) with increase age group, alcohol consumption, smoking in these two last groups. Besides these risks, an observational prospective study identified further stroke risks factors, namely coronary disease (CI: 0.67–0.98) (p <0.043) and heart failure (CI: 0.07–0.68) (p <0.009), a failure, which obviously leads to heart attack (Fekadu et al., 2019). Further, since stroke outcome varies with some key vascular risk factors like diabetes, analysis around this outcome should always adjust or control these factors (European stroke organization, 2017). Field (2018) additionally reported in the same sense that, provided the predictors correlation with the outcome, they are hierarchically added in a model by the enter method to prevent suppressor effects (pp. 398, 400). In addition to the predictors variables themselves, these four variables (demographic variable (age); behavioral risk factor (smoking and alcohol consumption); vascular risk factor (diabetes)) above reported to be the risk factors for stroke, are considered as confounders, and thus, adjusted or controlled for in the hierarchical logistic regression analysis of this research to ensure not just a more reliable result, but also to ensure that bias are minimized.

While the model fit is read from Nagelkerke's R^2, because this value does not only indicate the reduction of errors (Field, 2018, p. 903) or the variance of the dependent variables explained by the independent variable(s), but then again predicts the outcome too. Besides, a significant better prediction of the outcome is nevertheless given by the sequential addition of a predictor to the model, thus, chi-square value of the omnibus test (Field, 2018, p. 897). In addition, the exponential of b delivers information on the effect size, that could be interpreted as an odd ratio as well (Field, 2018, p. 904). According to Field (2018) significance is given by the Wald statistics p-values (p. 902). Furthermore, to find out whether the predictors are independent from each other, hence testing purposefully multicollinearity, a linear regression with same outcome and continuous metric predictors BMI, SES score is run, however only the VIF is read and considered from the result output (Field, 2018, pp. 913-914). As far as other aspect like interaction between the predictor's variables (BMI and SES score) is concerned, its handling, is discussed in Moderator effects section hereafter.

Table 6: Overview of the variables, level of measurements and reference group used in the statistics

Variable	Type of data in DEGS1	Modified level	Reference
Gender	Nominal dichotomous	-	-
SES	Ordinal (3 Levels)	Dummy coded	Middle SES
SES Score	Continuous metric	-	-
Blood pressure	Nominal dichotomous	-	-
BMI	Continuous metric	-	-
BMI categories as defined by the WHO	-	-	-

Stroke	Nominal dichotomous	-	
Age (confounding)	Continuous metric	-	-
Alcohol consumption (confounding)	Nominal dichotomous	-	-
Smoking (confounding)	Nominal dichotomous	-	-
Diabetes (confounding)	Nominal dichotomous	-	-

3.6.4 Moderator Effect

Moderators or effects modifiers are variables that affect either both the direction and strength of relationship between variables other in some cases might just affect the association strength. In other words, the effect of an independent variable on a dependent variable, depends on the expression of another variable called the moderator. Field (2018) defines these variables as those susceptible to affect an association between a predictor and an outcome variable (p. 484). Field (2018) even goes further and states that moderation happens under the condition that, in a regression analysis, the interaction resulting between a predictor and an effect modifier variable must be of a significant nature (p. 497). To include continuous predictors and moderators in a regression analysis, their means are being centered in advance in order to be able to confer the results a clearer and an explicit comprehension on one side and on the other side, because the variable BMI is redundant. Furthermore, the means are as well centered to still be able to meaningfully interpret the main effect, even when there is no interaction effect. Mean centered variables are obtained by deducting their respective average mean value from the corresponding individual value of the same variable (Field, 2018, p. 487).

In course of this study, an interaction is tested between the metric variables BMI mean centered, and SES score mean centered, then it is besides imperative to figure out by a validation or invalidation of the assumption whether their accordingly effects action are indeed and actually independent or rather combined.

This section gave a deep insight of the methodology particularly the different statistical analysis that are performed to be able to achieve the objectives and obtained scientific results. All obtained results from the data analysis are presented in the following section.

4 Results

In this section, the results of the analyses are detailly described and presented. At first, the descriptive statistics are illustrated and described in percentage and numbers where necessary. The core analysis for difference assessment follows the descriptive statistics and enough comprehensible results of crosstabulation and U Mann Whitney of the variables are presented in table forms. Next to the bivariate analysis, the main results of the hierarchical logistic regression are presented considering its adjustment for confounding variables and moderation effect. For better and rapid comprehension purposes, the results of the analysis are illustrated on bar graphs and pie chart and table as well.

4.1 Sample description

The sample contains n=7987 subjects of which the most represented gender is female (52.6%) and the less represented are the men (3789) with a percentage of 47.4%. Figure 7 shows the valid percentage of subject according to their gender included in the analysis.

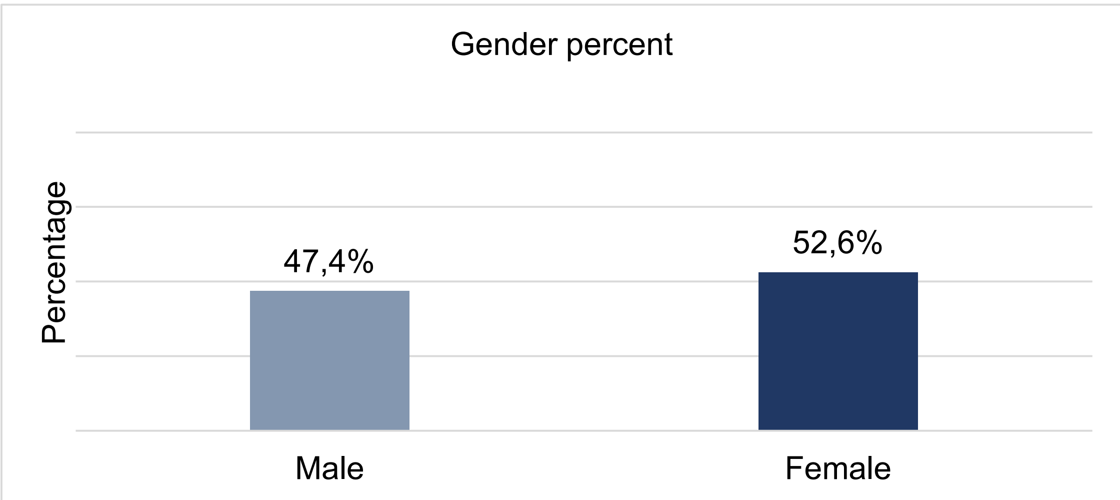


Figure 7: Valid percentage of subject according to their gender

The overwhelming majority (males: 2147; females: 2596) of the valid subjects come from a middle SES, while subjects from the low SES are less represented. Approximately 27% of the men have a high SES index compared to that of the women, which lies around 21%. Moreover, there are altogether 90 cases accounting for missing values which were left out of the analysis. Figure 8 shows the valid percentage of subject according to their SES.

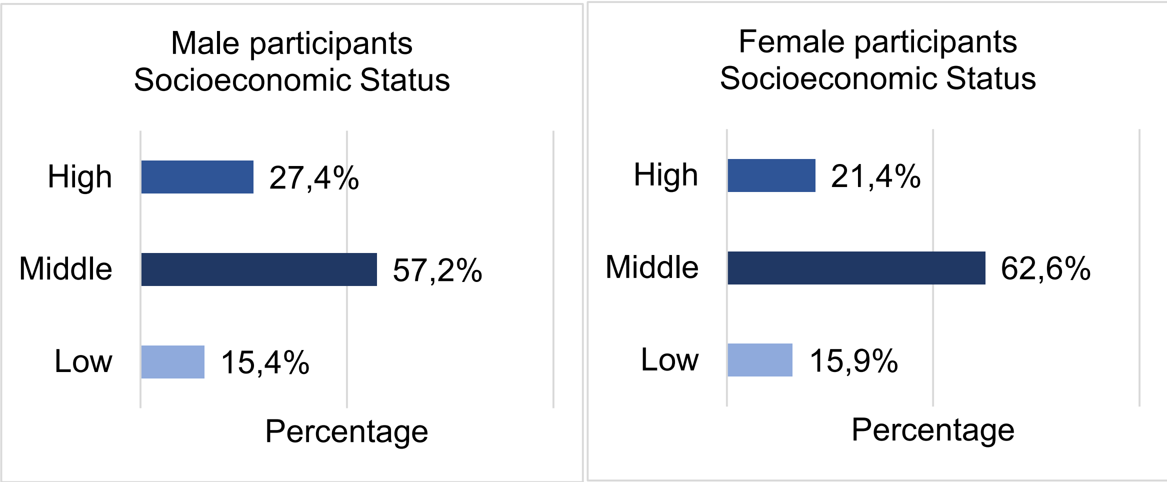


Figure 8: Valid percentage of subject according to their SES

Further descriptions concern the distribution of subject according to their respective BMI as defined by the WHO, and it is noticed that most of the male respondents (1503 and thus,44.3%) are predisposed to obesity against 26.6% of the women, whereas the percent of the both the obese females and obese males is the same (25%). However, the number of underweight subjects (both genders) is quit under 2%. From the variable BMI group, 922 cases had a missing value. On Figure 9, the valid percentage of subject´s BMI according to their BMI categories as defined by WHO can be observed.

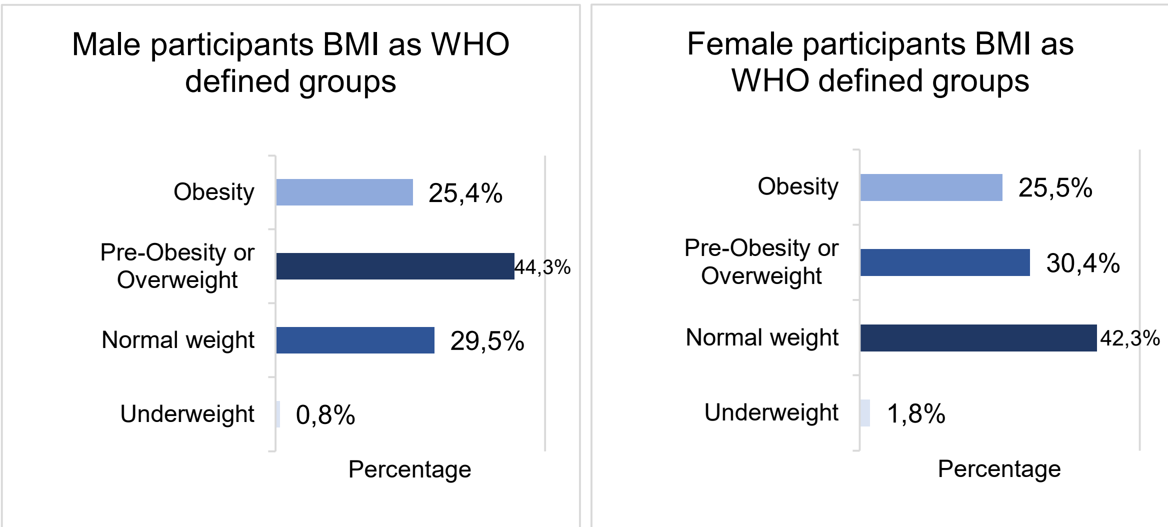


Figure 9: Valid percentage of subject´s BMI according to their BMI categories as defined by the WHO

Figure 10 gives information on how many subjects were medically diagnosed with high blood pressure and thus, more than 38% of the subjects were positively diagnosed i.e., the men suffer mostly from high blood pressure than the women do (41% and 36.4% respectively). However more than 61% of the subjects are not attained with the hypertension as the diseases was pictured out by the WHO.

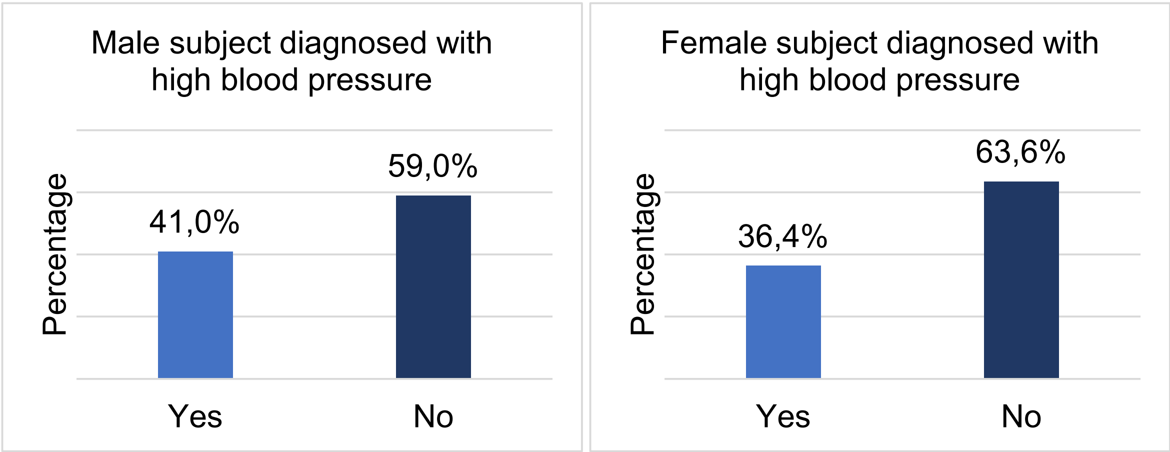


Figure 10: Valid percentage of subject diagnosed with high blood pressure

In the whole sample, 3.2% is the valid percent of stroke, of which the males are about two time more affected than the females (4.1 against 2,4), as illustrated in Figure 11. Those who reported not to know about their health status as far as stroke is concerned were recorded as missing and the overall missing cases were 1019 for the males and 1125 for the females.

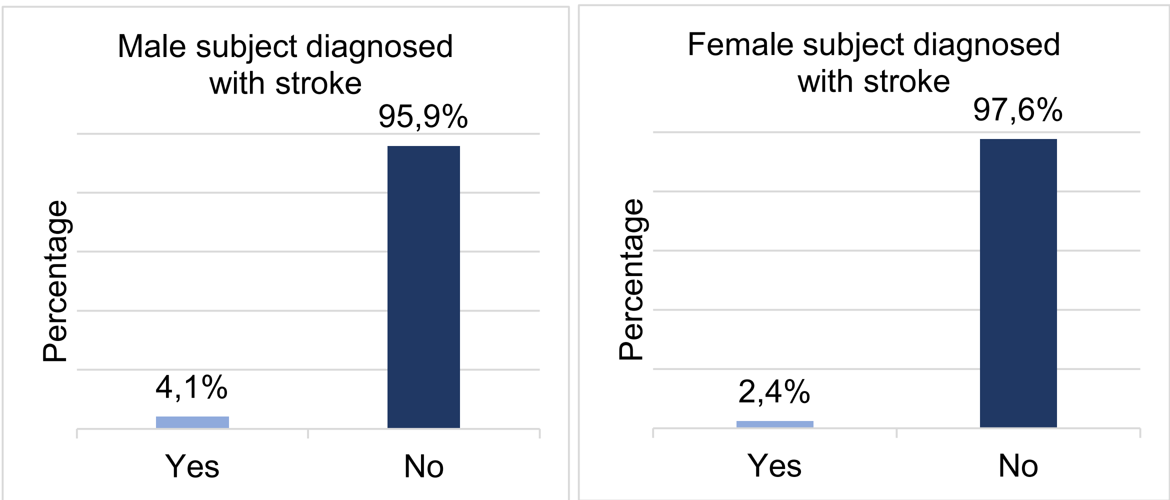


Figure 11: Valid percentage of subject diagnosed with stroke

Considering the total sample, an exploring analysis crossing stroke against hypertension shows that the valid percent of males living with hypertension and suffer from stroke is slightly inferior to that of females living with hypertension and suffering from stroke like it can be observed from the pie chart Figure 12.

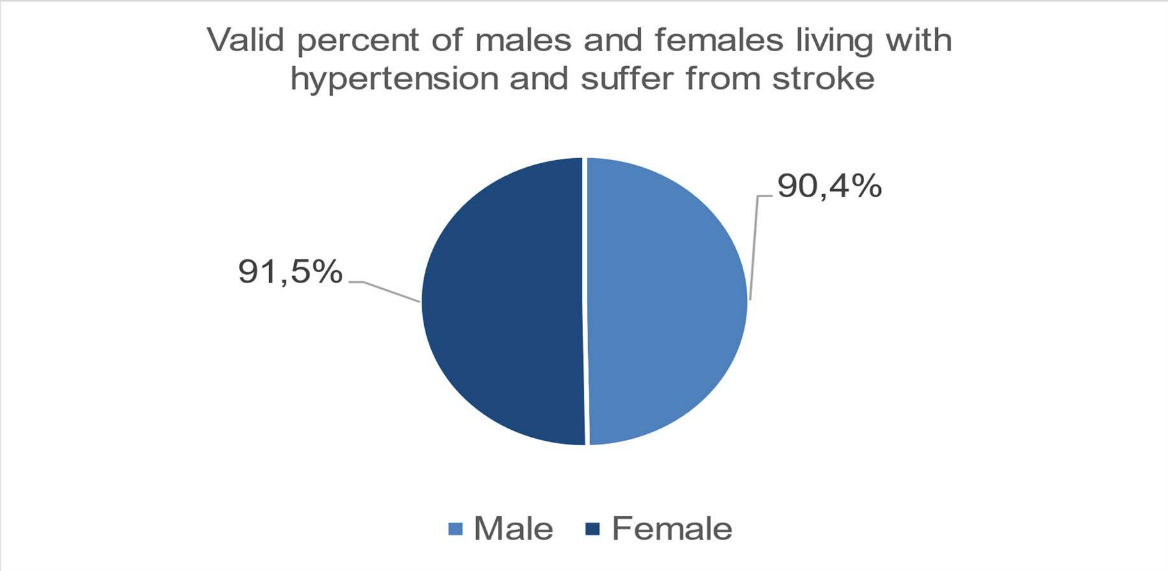


Figure 12: Valid percentage of males and females living with hypertension and suffering from stroke

Table 7 hereafter gives an overview of the mean, standard error, standard deviation, variance as well as minimum and maximum of the continuous variables. The values here are rounded to the nearest tenth.

To prevent distorting effects and for reasons of better visualization of the gender differences in the description of BMI and SES score, the data split by gender gave the results shown in Table 7 below. Here, the male SES score mean and standard deviation ($\bar{x}$ = 11.68; SD=3.81) does not differs that much from that of female ($\bar{x}$ = 11.33; SD=3.56). However, the SES score variance is higher in men (Var = 14.88) than in women (Var = 12.61). Both the male BMI mean and standard deviation ($\bar{x}$ = 27.37; SD = 4.44) are slightly higher than that of the females ($\bar{x}$ = 26.69; SD = 5.57), however the BMI variance is much higher in women (Var = 30.98) than in men (Var = 19.74). Furthermore, the SES score ranges from 3.00 to 21.00 for both males and females whereas the BMI ranges from approximately 15.79 to 56.85 kg/m² for both males and females. The missing values for the variables: score and BMI are 437 for men and 575 for females.

Table 7: Descriptive statistics of the continuous variables

Gender	Variables	N	Mean	SE	SD	Variance	Min	Max
Male	SES Score	3751	11.683	0.062	3.806	14.88	3.00	21.00
	BMI	3390	27.374	0.076	4.443	19.742	15.794	56.853
Female	SES Score	4146	11.328	0.055	3.551	12.606	3.00	21.00
	BMI	3675	26.689	0.092	5.566	30.980	15.988	54.5990

Table 8: Kolmogorov-Smirnov test of normality

Gender		Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Male	SES-Score	0.070	3366	<0.001	0.977	3366	<0.001
	BMI	0.051	3366	<0.001	0.961	3366	<0.001
Female	SES-Score	0.040	3651	<0.001	0.988	3651	<0.001
	BMI	0.081	3651	<0.001	0.940	3651	<0.001

The Kolmogorov-Smirnov test for normal distribution (Table 8) shows a highly significant results in all the two metric variables and in both the male and the female group. This result shows strong evidence against null hypothesis that, both SES score BMI are normal distributed. SES score normality test results for men are (D (3366) = 0.070; p<0.001) and for women (D (3651) = 0.040; p<0.001), while BMI normality test results for the males (D (3366) = 0.051; p<0.001) and for the females are (D (3651) = 0.081; p<0.001).

These results demonstrate that the variables are not normally distributed. To gain better insight into their distribution, Figure 13 and Figure 14 hereafter show the histograms of the variables with their distinct distribution.

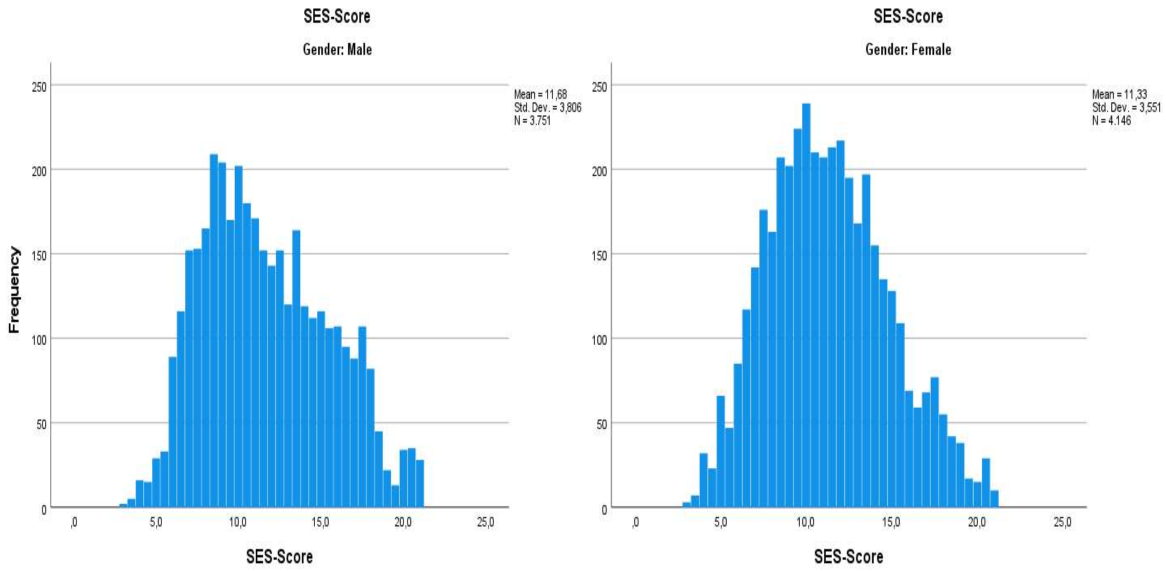


Figure 13: Histogram showing Socioeconomic Score distribution

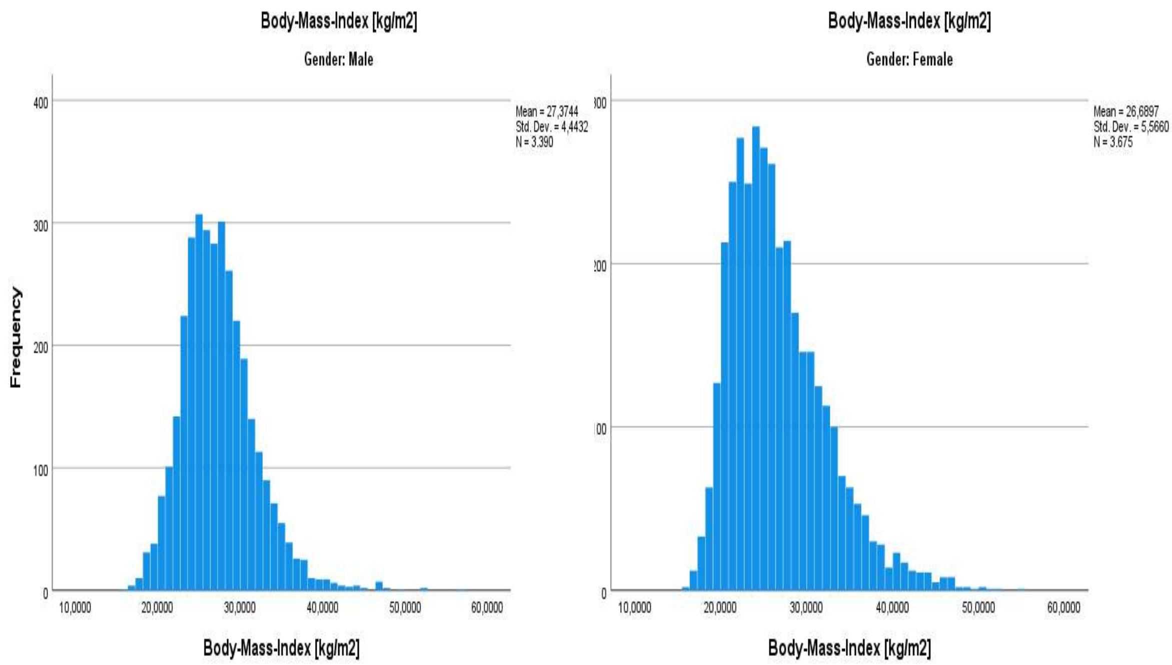


Figure 14: Histogram showing Body Mass Index distribution

4.2 Difference assessments

In this section, the different assessments results for the analysis of the variables are presented.

According to the crosstabulation results (Figure 15) to assess differences, there is a gender difference in stroke. More than 60% of the men are exposed to stroke compared to an exposure percent of just almost 40% for the women. This implies that men have approximately 20% more chance of suffering from stroke than women do. Furthermore, 4,1% of the men suffer from stroke in contrast to 2,4% of the women.

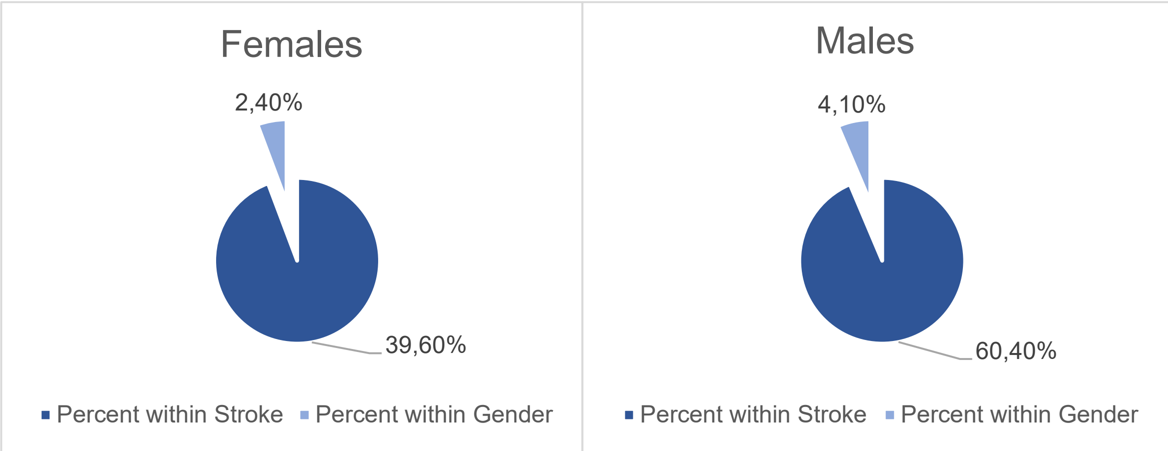


Figure 15: Percentage showing gender difference in stroke

When a look is taken at the Figure 16 below, it is noticed that 81,3% of male subjects living with high blood pressure are exposed to stroke compared to 86% of the female subjects. In other words, women have an approximate 5% much more chances to be exposed to stroke than the men do. Moreover, almost 7% of the men living with hypertension suffer from stroke, while exactly 1.5% of the males living without the disease suffer from stroke. In contrast to the males, 4.6% of the females living with hypertension suffer from stroke, whereas just 0.6% of the females living without the disease suffer from stroke. Therefore, stroke prevalence in subjects living with stroke is said here to be distinctly higher in men than in women.

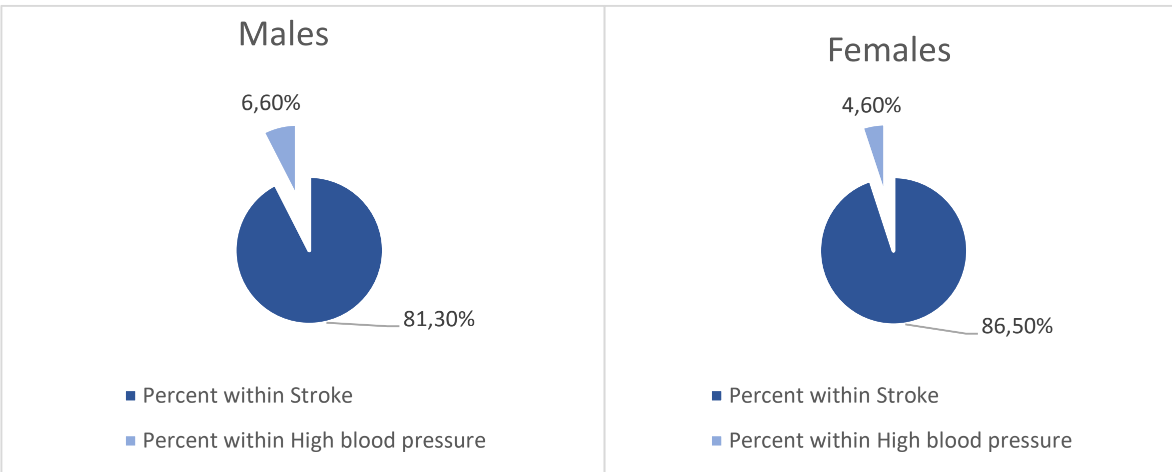


Figure 16: Percentage showing high blood pressure difference in stroke

U Mann Whitney assessing difference shows in its output (Table 9) that, there is a highly statistical significance evidence against the H_1, thus there is a SES score difference in stroke outcome for males (U = 94526.500; p=0.009) and females (U = 54361.500; p<0.001) as well. Regarding BMI difference in stroke, the female group is the only group presenting a statistically significant difference (Mann Whitney, U = 60023,500, p<0.001). The male group in contrast do not (Mann Whitney, U = 101403.000, p=0.115).

Table 9: U Mann Whitney SES score and BMI difference in stroke

Gender	Variables	Mann-Whitney U	Wilcoxon W	Z - Score	P - Value (2-tailed)
Male	SES Score	94526.500	99182.500	- 2.598	0.009
	BMI	101403.000	2826348.000	- 1.578	0.115
Female	SES Score	54361.500	56506.500	- 4.915	<0.001
	BMI	60023.500	3438723.500	- 3.991	<0.001

4.3 Bivariate analysis

Even though the metric variables are not normally distributed, the correlations between the BMI and stroke, SES score and stroke are still determined by point biserial correlation in disfavor of the non-parametric test, Spearman´s rho. Firstly, this is because the point biserial is a robust test; secondly, due to the fact that the test variable of this study is measured on a nominal binary level, in such a way that the most important assumption for this very test is met; thirdly, it has to be underlined that the assumption of no outliers is met for both males and females. However, some males and females BMI values are to be seen on the top whisker of the boxplots (Figure 17), thus violating the assumption of no outliers; lastly, the study include a large sample size enough. In the same vein, DeCoster, J. (2004) moreover states that Spearman correlation is a less powerful measure used to capture association, so that Pearson correlation is commonly choose, even when the variables considered are moderately non normal distributed (p.29).

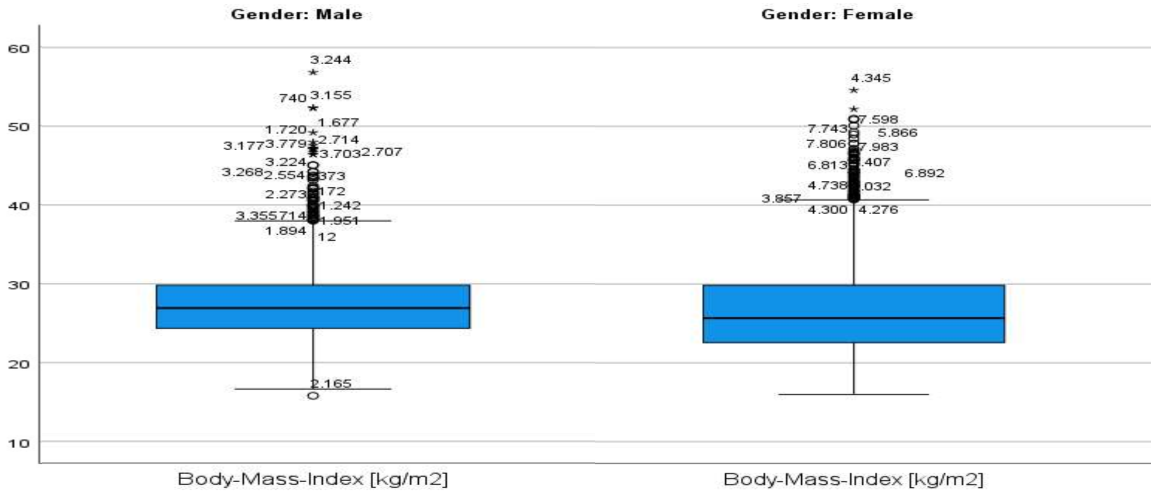


Figure 17: Boxplots showing BMI outliers for both genders

As it is visible below in Table 10, SES as a score is highly positively significantly associated with stroke in males and females as well (males: r = 0.054; p=0.008; n=2430; females: r = 0.094; p<0.001; n=2664), meaning that whenever SES score increases, it causes alongside a significant increase in stroke prevalence. Still on Table 10, it is further besides to notice that BMI is negatively associated with stroke in men and women as well. Nevertheless, BMI is weakly and not significantly associated to stroke for males (r = - 0.024; p=0.241; n=2430), whereas shows a highly significant and strong negative association to stroke for females (r = -0.070; p<0.001; n=2664). That is, a rise in BMI value, induces a significantly decrease in stroke prevalence for female, but a non-significant decrease of stroke prevalence for males.

Table 10: Point-Biserial correlation between stroke, BMI and SES score split by gender

Gender	Variables	Pearson Correlation (r)	P - Value (2-tailed)
Male	SES Score	0.054	0.008
	BMI	- 0.024	0.241
Female	SES Score	0.094	< 0.001
	BMI	- 0.070	< 0.001

As far as that association between independent categorical variables and stroke is concerned, chi square test reveal that, 0 cells (0.0%) have expected counts less than 5 and that the minimum expected count for gender is 88.63 and for high blood pressure sorted by gender 55.65, thus indicating that the sample size assumption for this test were not violated. Gender is significantly highly associated with stroke. The Chi-square results shows that this association is significant, X^2 (1) = 13.154; p< 0.001, even if the effect size read from Phi output is not existent according to Cohen, p<0.001 and effect size=0.047. As Table 11 reports, after gender was sorted, Chi-Square test output shows that stroke depends on high blood pressure for both males and females. The results of this association are highly significant too (males: Chi-Square, X^2 (1) = 44.714; p< 0.001; females: Chi-Square, X^2 (1) = 52.019; p< 0.001). This time around the association between hypertension and stroke hat a statistically significant effect size slightly higher that than that for the association between gender and stroke (Phi=0.128; p < 0.001), even if both associations are still classified under the category no effect. It should be further remembered that the effect size description varies from one source to another, for instance, Sensu Hattie (2009) would have interpreted the effect-sizes of these two Chi-Square results as "*developmental effects*".

Table 11: Chi-Square Tests showing correlation between stroke and high blood pressure split by gender

Gender	Pearson Chi-Square (X^2)	df	P - Value (2-tailed)
Male	44.714	1	< 0.001
Female	52.019	1	< 0.001
Male and Female	13.154	1	< 0.001

4.4 Binary logistic regression on stroke

The hierarchical analysis comprises three models, in which the variables are subsequently added by the enter method and with respect to the categories i.e., in model *I*, depending on whether they are confounders, in the model *II* depending on whether they are the core predictors of stroke and thus, the core assessment of this study, or in model *III*, depending on if they are moderators. Withing each model, the independent variables are added following an increasing order of their respective correlation coefficient with the outcome i.e., from the highly correlated to the lowest. Before gender splitting, a binary regression (first) is run with all confounders, predictors, and moderators, however only the result coefficient of the predictor variable gender is read from the analysis, so that a second binary regression analysis after gender splitting excluded the variable (gender) from the analysis. Moreover, multicollinearity for the variables BMI and SES score is tested with linear regression equation and the VIF results (Appendix VI) revealed an a no multicollinearity between the variables, since the VIF value for males for both variables is 1.012 and for females 1.063. These values are distinctly lying under the threshold 10, a threshold above which, an existing problematic multi-collinearity is characterized (Field, 2013, pp. 324-325).

4.4.1 Binary logistic regression results, males compared to female participants

The regression on stroke included more females as male (2535 compared to 2306) of which 4.1% of the latter suffer from stroke in contrast to 2.1% of the former.

The regression analysis was executed several times, but with different settings at each execution like amongst other: the inclusion of gender as predictor or the shifting of the cut off value to detect changes in the predicted probabilities.

Looking at the Omnibus tests of the model summary in Table 12 and Table 13 below, it can be observed that the models *I* and *II* for both genders are highly significant with each time a p-value less <0.001. In contrast to the Omnibus tests results, apart from the model *I* for the females, the Hosmer and Lemeshow test shows a non-significant result in all three models for both genders, thus indicating that there would be no difference between observed and expected cases and hence, the data appropriately fit the goodness of the models. The Hosmer and Lemeshow test result for females in the model *I* exploring whether the predicted and the observed probabilities are the same is significant (X^2 (8) = 15.645; p = 0.048), thus, an indication of the confounders not fitting into the model. It should be moreover noticed that some expected cases in the contingency table report a value lower than 5, therefore making the above interpretation of Hosmer Lemeshow difficult.

Table 12: Model summary for logistic regression on stroke for males

Models	Variable	Nagelkerke´s R^2	Omnibus tests		Hosmer and Lemeshow Test	
			Chi-Square (df)	p-Value	Chi-Square (df)	p-Value
I	Age Alcohol consumption Cigarette consumption Diabetes	0.096	63.623 (4)	<0.001	9.961 (8)	0.268
II	High blood pressure Low SES High SES BMI	0.133	25.732 (4)	<0.001	7.551 (8)	0.479
III	Mod BMI SES	0.134	0.490 (1)	0.484	10.978 (8)	0.203

Table 13: Model summary for logistic regression on stroke for females

Models	Variable	Nagelkerke´s R^2	Omnibus tests		Hosmer and Lemeshow Test	
			Chi-Square (df)	p-Value	Chi-Square (df)	p-Value
I	Age Alcohol consumption Cigarette consumption Diabetes	0.127	61.414 (4)	<0.001	15,645 (8)	0.048
II	High blood pressure Low SES High SES BMI	0.176	24.162 (4)	<0.001	7,204 (8)	0.515
III	Mod BMI SES	0.179	1.477 (1)	0.224	6.188 (8)	0.626

Further lectures on the model *I*, assessing and controlling for confounders (Table 14 and Table 15) however indicate, that age is a confounder for both males and females (males: X^2 (1) = 29.290; p<0.001 and females: X^2 (1) = 9.268; p = 0.002) and further show that the behavioral risk alcohol consumption is a confounder for females (X^2 (1) = 4.140; p = 0.042), but not for male. Where X^2 represents the Wald value to be read on in the model estimates results.

After controlling for possible confounders Table 12 and Table 13 further show that, model *II* significantly predicts stroke (males: X^2 (4) = 25.732; p < 0.001; Nagelkercke R^2 = 13.3%; females: X^2 (4) = 24.162; p < 0.001; Nagelkercke R^2 = 17.6%). This implies that 13.3% (males) and 17.6% (females) of the variance in the equation of the model is explained by the predictors or independent variables BMI, SES, and hypertension. Hence, the variable added in the equation, fit mostly the women model as compared to males, since it has a much higher variance accounted for by the independent variables.

Overall, to be read in males and females summaries is that, the predictors high blood pressure, low SES as compared to middle SES have a significant influence on stroke, whereas BMI, high SES as compared to middle SES have a non-statistically significant influence on the outcome. Further lectures reveals that, if a male subject is being diagnosed with hypertension, he would be 0.325 times (b = -1.124; SE = 0.287; p < 0.001) less likely to suffer from stroke compared to a male subject who does not suffer from the disease (95% CI: 0.185 -0.570), whereas, a female subject diagnosed with hypertension is 0.231 times (b = -1.464; SE = 0.431; p=0.001) less likely exposed to stroke than a woman, who does not suffer from stroke (95% CI: 0.099 - 0.538). Therefore, in a hypertension context, men are 0.094 times less likely to suffer from stroke than women are and reversely women are 0.094 times more likely to suffer from stroke than men are, hence high blood pressure could be considered in this context as a protective factor. The significant odds of stroke for a man originating from a low SES as compared to a man originating from a middle SES is 0.549 times lower (b = -0.600; SE = 0.277; p = 0.030; 95% CI:0.319 - 0.944), while the significant odds of stroke for a woman originating from a low SES as compared to a woman originating from a middle SES is 0.509 time lower (b = -0.675; SE = 0.322; p = 0.036; 95% CI:0.271 - 0.956). Each time a person originates from a high SES, he or she is highly exposed to stroke, but in a non-statistically significant manner (males: b = 0.324; Exp(b) = 1.383; SE = 0.286; p = 0.257; 95% CI:0.789 - 2.422 and females: b = 0.997; Exp(b) = 2.710; SE = 0.617; p = 0.106; 95% CI:0.809 - 9.082). After centering BMI (redundant variable) in the model of this study, it did not predicted stroke neither for males (b = 0.032; Exp(b) = 1.033; SE = 0.029; p = 0.273; 95% CI:0.975 - 1.094) nor for females (b = -0.019; Exp(b) = 0.981; SE = 0.029; p = 0.511; 95% CI: 0.927 - 1.039).

The interaction and thus the moderation effect between BMI and SES score assessed in model *III* shows a non-significant omnibus test of the model fit and therefore a non-significant result both for males and females in the results equation, signifying that, there is not a statistically moderation effect between the two variables that one could rely on as defined by Field in 2018, when he states that moderation occurs if in a regression analysis, the interaction resulting between a predictor and an effect modifier variable is of a significant nature (p. 497), (males: b = 0.005; Exp(b) = 1.005; SE = 0.007; p = 0.486; 95% CI:0.991 - 1.019 and females: b = -0.010; Exp(b) = 0.990; SE = 0.008; p = 0.217; 95% CI:0.974 - 1.006). The addition of the interaction between BMI and SES score as model 3 to the whole model studied, did not accounted for any extra variance change of the main model (model *II*) for neither male nor female group.

Table 14: Model estimates for regression on stroke in males

Models	Variables	B	S.E.	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
I	Age	-0.072	0.013	1	<0.001	0.931	0.907	0.955
	Alcohol consumption	-0.018	0.232	1	0.939	0.982	0.624	1.547
	Cigarette consumption	-0.373	0.328	1	0.255	0.689	0.362	1.309
	Diabetes	-0.361	0.270	1	0.180	0.697	0.411	1.182
II	High blood pressure	-1.124	0.287	1	<0.001	0.325	0.185	0.570

	Low SES	-0.600	0.277	1	0.030	0.549	0.319	0.944
	High SES	0.324	0.286	1	0.257	1.383	0.789	2.422
	BMI	0.032	0.029	1	0.273	1.033	0.975	1.094
III	Mod BMI SES	0.005	0.007	1	0.486	1.005	0.991	1.019

Table 15: Model estimates for regression on stroke in females

Models	Variables	B	S.E.	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
I	Age	-0.052	0.017	1	0.002	0.949	0.918	0.982
	Alcohol consumption	1.231	0.605	1	0.042	3.426	1.046	11.218
	Cigarette consumption	0.570	0.620	1	0.358	1.768	0.525	5.957
	Diabetes	-0.660	0.331	1	0.046	0.517	0.270	0.989
II	High blood pressure	-1.464	0.431	1	0.001	0.231	0.099	0.538
	Low SES	-0.675	0.322	1	0.036	0.509	0.271	0.956
	High SES	0.997	0.617	1	0.106	2.710	0.809	9.082
	BMI	-0.019	0.029	1	0.511	0.981	0.927	1.039
III	Mod BMI SES	-0.010	0.008	1	0.217	0.990	0.974	1.006

The model *II* classification is more than 70% completed (males: 96%; females: 97.8%), however not optimally, then the percent of correct predicted values in each characteristic was not above 50% (Schillmöller, 2019), but rather exactly 100% for the observed cases not diagnosed with stroke and 0% for the observed stroke cases. The classification unfortunately predicted no stroke cases for males and females as well, hence sensitivity 0%. Nevertheless, a false negative rate of around 4% and a true negative rate of approximately 96% were predicted for males compared to 2.2% and 97.8% for the respective rates for females. Overall, the specificity is more expressed than the sensitivity. Thus, the sensitivity of a test being its ability to capture true positive cases, the cut off value set at 0.5 was too high for the test to capture true positive. In fact, a shifting of this cut off value, in the sense of increasing it or decreasing it would be at the expense of the sensitivity for the former scenario or at the expense of the specificity for the latter scenario. Indeed, for the purpose of more comprehension, the cut off is shifted below and above 0.5 and a consequent change (i.e., some few stroke cases are capture) is observed in the classification of the equation, whenever this value is set at 0.8 or 0.9. If not in the classification tables, no other further changes are observed either in the model summary nor variable in the equation results.

A separate analysis (Table 16; Appendix II) included the variable gender with other predictors, nevertheless with only objective of assessing the predicting character of this former variable on stroke and then, this same variable (gender) was excluded from all other models, because the same variable served as basis on which the analysis was split to enable the analysis to be run separately for males and females. The omnibus results indicate that, gender is a highly significantly predictor for stroke and the model estimates that the data fit well the model, further, the model including gender account for 15.1% of the variance of the stroke. The odds of an individual to be diagnosed for stroke are 0.524 times lower for males as compared to females (b = -0.646; SE = 0.182; $p < 0.001$; 95% CI: 0.367 - 0.748). That is, an individual with the gender male are 0.524 times less likely to suffer from stroke as an individual with the gender female. In the next section, the bivariate, regression analysis results obtained in course of the statistical analysis are discussed.

Table 16: Model estimates for regression on stroke including the variable gender

Model	Variables	B	S.E.	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
II	Gender	-0.646	0.182	1	<0.001	0.524	0.367	0.748
	High blood pressure	-1.253	0.238	1	<0.001	0.286	0.179	0.455
	Low SES	-0.631	0.209	1	0.003	0.532	0.353	0.802
	High SES	0.479	0.254	1	0.059	1,614	0.981	2.654
	BMI	0.008	0.020	1	0.708	1.008	0.968	1.049

5 Discussion

The descriptive statistics shows that more women were included in this study as men were. Most of the individuals came from a middle social economic neighborhood, but subjects with low and high socioeconomic status were highly represented as well. Women seem to have a better control of their own weight as the men do, since most of them have normal weight. The biggest proportion of the males in contrast are represented in the overweight category (as defined by the WHO). The proportion of underweight individuals in the sample is quasi inexistant, approximately 1% for the males and 2% for the females. Indeed, precisely in a context of hypertension and according to the statistics like showed in Figure 15, the women are far ahead of the males, when it comes to the chance of exposure to the outcome stroke or otherwise said, women are most exposed or the gender more likely to suffer from stroke as the men are. However, when a look is taken at Figure 12, it is rather unfortunate that the females are the less attained, even though their percentage of exposition to the outcome in a hypertension context is clearly higher. The American Stroke Association states that women in general are more at risk of stroke than men do, und this might be due to several reasons. On one hand, the risk of the outcome in pregnancy time is 21 per 100 000 females and this risk increases even further during the third trimester and post-partum period. On the other hand, women under birth control pills, which simultaneously consume cigarette are much more exposed to stroke (American Stroke Association, 2018).

The relationship between gender and stroke is highly significant, even though it has a quasi-inexistent effect size as per Cohen effect size classification. Furthermore, SES in general, as a score is highly positively associated with stroke for both male and female. This positive relationship contradicts the inverse correlation found by Busch et al., (2013). BMI in contrast, is admittedly associated with stroke for equally for men as for women, however, the nature of this relationship is negative in both genders; weak and not significant for males but rather strong and highly significant for females. Additionally, after splitting the sample by gender, Chi-Square test output is highly statistically significant and shows that stroke depends on high blood pressure for both males and females. This Chi-Square results is in phase with the findings published in 1990 and 2010 after a research conducted on the association between hypertension as independent and stroke as dependent variable, stating that in the first 48 hours or 2 days of stroke onset, 80% of the outcome is associated with high blood pressure (Dunbabin et al., 1990; O´Donnell et al., 2010, as cited in Furlan et al., 2018). Though the effect size here once more is quasi-inexistent, it is nevertheless slightly higher than that for the association between gender and stroke presented previously.

There was an association between stroke and hypertension as revealed by the Chi-Square test of association output as seen in section 4.3, however the nature of the association could not be determined by this test. A better insight in the association nature could be established by a hierarchical binary regression model. The logistic regression model controlling for age, diabetes, alcohol, and cigarettes consumption revealed that: age is a confounder equally for male as for female, however, the behavioral risk, thus alcohol consumption is a confounder just for the female, however not for the male. Next to the confounding adjustment moreover, the predictors: gender, SES, high blood pressure and BMI included in the equation has fitted the model in a such a way that, its summary shows that these independent variables explained 15.1% of the variance of the dependent or outcome variable (stroke). When the model rather excluded the predictor gender (since it served as basis on which the analysis was split), its summary shows that the rest of the independent variables obviously explained less and, thus rather 13.3% of variance accounting for the outcome stroke in males, whereas its summary explained more the females in contrast, thus, a variance of 17.6% accounting for the outcome variable. Further results to be considered is that, the odds a male subject to suffer from stroke is 0.524 times lower than that for a female subject (CI:0.366 - 0.746). This regression results implies that women are more exposed to the outcome than males are. It is however rather unfortunate, because according to Figure 11, males are the sex most attained with stroke, perhaps for pregnancy and birth controls pills reasons liked expressed by the American Stroke association in 2018 here above. In the same vein, compared to individuals living without high blood pressure, hypertensive women are furthermore according to Table 15 slightly more likely to suffer from stroke than hypertensive men are (Table 14). Stroke affects both the gender male and female even though there is a distinguishable aspect between the incidence rates and outcomes of these genders i.e., age-specific stroke rates is said to be higher in men than in women, though female individuals further have to deal with more substantial events due to stroke because, they live more longer, and thus have a much more long-life expectancy. As a result, their stroke incidence at older ages is therefore said to be indeed undoubtedly higher (Reeve et al., 2008, as cited in Hiraga, 2017).

While the Odds Ratio (OR) of males living with high blood pressure is 0.325 times less than that of males living without the disease, females living with high blood pressure have an OR of 0.231 times less than that of females living without the disease. Overall, high blood pressure could be considered in this context to be a protective factor against stroke, since it is associated with a lower likelihood of a negative outcomes (then individuals living with high blood pressure have a weaker chance of developing stroke) or hypertension might be understood as a "*positive countering events"* (Substance Abuse and Mental Health Services Administration, n.d., p.1). The odds of females coming from a low SES neighborhood is 0.509 time significantly less than that of females originating from a middle SES neighborhood and in the same manner, the odds of men coming from a low SES neighborhood is 0.549 time significantly less than for those originating from a middle SES neighborhood. That implies that, living in a low socio-economic society predisposes both the men and the women to a lower significant chance of developing stroke as compared to living in a middle socio-economic society, though, when an individual comes from a high socio-economic society, he or she is highly (1.383 times more for the males and 2.710 times more for the females) exposed to stroke, but in a non-statistically significant manner. This result rejoins the findings of Dunbabin et al. (1990); O´Donnell et al. (2010), as they found out that stroke changes in a directly proportional way with SES i.e. the prevalence of the disease or an individual chance to get stroke incresases with an increase in SES level. However, this study findings contradicts what was found by Busch et al. (2013) as after carrying out a study, an inversely proportional association (inverse social gradient) between the two variables was found, thus rejecting the theory that stroke risk decreases with an increase in SES level. To sum up, the females are almost similarly significantly exposed in the low SES neighborhood (OR males: 0.549 and OR females: 0.509), however differently non-significantly exposed in the high SES neighborhood (OR *males: 1.383 and OR females: 2.710).* After centering BMI, this variable did not significantly predict stroke, neither for females nor for males.

Furthermore, after searching for an eventual interaction between BMI and SES score, the results revealed that there was no interaction and thus, no subsequent moderator effect between BMI and SES score bearing in mind that the addition of the corresponding interaction into the model neither was significant nor changed the Nagelkercke R^2 value, indicating that the variance of equation in the main model of the study remained constant.

Since this study methodology is mainly based on the analysis of secondary data, and even though secondary data analysis have some advantages, the interpreted of the results were done with caution because the method presents some limitations that should be considered as well.

5.1 Strengths and Limitations of the Study

"In a time where vast amounts of data are being collected and archived by researchers all over the world, the practicality of utilizing existing data for research is becoming more prevalent" (Andrews et al., 2012; Schutt, 2011; Smith, 2008; Smith et al., 2011, as cited in Johnston, 2014). Secondary data analysis could refer to a research or study, in which, in course of its methodology and accordingly analysis, data collected for primary purposes by an organization, a research institute, a university etc., are reused by a researcher to answer a particular research question, to address a particular problem or to fill a research gap. The usage of existing data provides a researcher with advantages and disadvantages as well.

5.1.1 Limitations of the Study

According to Columb & Atkinson (2016) type I errors are caused by uncontrolled confounding influences (p. 159). Following this reasoning, the control for confounders in this study effectively reduces the chance of type I error occurring, however the null hypothesis could have been incorrectly rejected due to the too large sample size. As a result, a 5% chance of a type I error might still have been existent and thus the assumption of an absolute null type I error would not have been guarantee or met.

Like any other research type, the current secondary data analysis has its own defined limits. An obvious disadvantage is the conceptualization of the DEGS itself. Even though it might have been broadly explained according to the RKI comprehension, it may not be possible in the context of this study to have a deep insight understanding of the how and the why (the data collected, the overall data collection process, the level of measurements of these data etc.) certain types of information was collected. Hence considerable time, efforts was invested trying to comprehend of what nature and structure the DEGS1 data set was. Additionally, the justification and explanation of the theoretical and analytical approaches within the frame of this study accounted for these efforts. Further, most measurements in the data set are repeatedly abbreviated, perhaps due to the RKI designed project setting to either serve multiple purposes or perhaps to be able to support a multidisciplinary team invested in the coordination of the institute's activities.

5.1.2 Strengths of the Study

Since data are collected in advance and are available in a dataset, the paramount strengths that could be directly linked with secondary analysis are cost-effectiveness and convenience that the secondary data availability provides the researcher with (Dale et al., 1988; Glaser, 1962; Smith, 2008, as cited in Johnston, 2014). This study gained access to DEGS from the RKI with all the advantages it may have including the utilization of high quality data, and the high level of representativity of the target population particularly the large sample size permitting a much more greater validity and much more generalizability of the findings (Smith, 2008; Smith et al., 2011, as cited in Johnston, 2014). Analyzing secondary data allowed this research to use quality data that could not be obtained in one way or the other on an individual level. A further strength is that, secondary analysis saves time as this study did not needed a particular design or any new set of data collection. The analysis of data collected by another party in this research is a learning process during which skills concerning the how or the management of raw materials to make incredible contributions to the scientific literature were developed. Furthermore, alternative, null hypothesis and accordingly discussion regarded from another perspective, completely different from what could have been researched and theorized by RKI were able to be tested in different setting in the context of this research.

5.2 Outlook for Action and Potential Future Research Directions

Based on the above results, the focus of interventions should be mostly put-on socioeconomic factors and blood pressure. Since the profession is a strong component of SES, action should then be oriented in this direction. Thus, companies should not only invest part of their capital for the professional training of its employees, but above all should spend money on the health education of their employees to promote this latter health status. Creating a particular department and hiring health personnel which could aid the employees to gain more Knowledge about health behaviors in the occupational field and as well help the workers to gain a deeper insight, up to date information on the most common diseases related to their workplace and were applicable, be informed on how to counteract these. The fact that individuals from high SES are more predisposed to stroke mean neither that there is a limitation to access to healthcare since the health system in Germany is based on the solidarity principle, nor that is an issue of lack of education since the more exposed are those in procession of a high degree level of education. It should be by the way recalled here that, SES index or score was derived from the participants different aspects of education, profession, and income, where most of the individuals suffering from stroke are those originating from a high SES, it would therefore imply that the reasons for the disease predisposition should be search elsewhere. Perhaps other factors (which could be confounders) like for instance stress, burnout or even too much ambition which might lead to putting oneself under high pressure conditions, could be at the origin of this predisposition and that should be taken into consideration by actors of the healthcare policy too. In this same vain, Salvagioni et al furthermore conducted a systematic review while following the guidelines of the Transparent Reporting of Systematic Reviews and Meta-Analyses in their methodology and came to the results that cardiovascular diseases including stroke, musculoskeletal pain, depressive symptoms, psychotropic and antidepressant treatment, job dissatisfaction and absenteeism are consistent effects of burnout (Salvagioni et al., 2017).

Besides, through sensibilization campaigns, online courses, notification by the health insurances or during hospitals control routines, awareness should be raised on blood pressure too. Indeed, it is well explicitly detailed and emphasized in the background that high blood pressure is a risk factor for stroke, however the population still have the perception or the consequent understanding that having no risk factors means not to be exposed at all and that is where the results of this study come into action by illustrating the opposite. This contradiction could be far more true or valid as well, then high blood pressure might be a protective factor too as established by the above results. Therefore, it is important to let people know that, having a risk factor or being exempt of whatever risk does not necessarily change something on the probabilities of developing stroke or other severe diseases, thought rather a non-exposed might be equally severely attained than an exposed could be.

The outlook for actions recommending some eventual initiative that could be undertaken regarding the findings of this study would be followed by the outlook for potential further directions, which could be applied to carry further research on stroke.

This study suggests several interesting additional avenues for future research. On one side, the findings here point to the need to further explore the possibility of an association between stress, burnout, hierarchical relationship between employees and their superior in company, and the prevalence of stroke in Germany. On the other side, it would be interesting to study the predicting character of these independent variables for stroke outcome. It further proved that stroke has a greater effect on women than men due to the experience of a much more events frequency compared to men on one hand and since females are less likely to recover from the disease on the other hand. Indeed, age-specific stroke rates are far ahead in men, however, because of women longer life expectancy and much higher incidence at older ages, they have more tendency to develop stroke events than men do (J Reeves et al., 2011, p.1). Part of these affirmations is in phase with the findings of this study, which besides were justified by the American Stroke Association (2018) as being due to a link with different phases of pregnancy and thus rendering it imperative that future research should focus on the other potential influencing factors of the personality, like for instance neuroticism, while correlating stroke in women and the other variables listed here above or any other variables that might have shown a difference between the gender.

6 Conclusion

This research used the data collected by the RKI within the framework of the DEGS study conducted between 2008 and 2011 to achieve it main aim and the various related objectives. This main aim formulated at the beginning of this study was to find out the association existing between independent variables: gender, SES, blood pressure, BMI, and the dependent variables stroke; and thereafter to assess whether including these former predictors variables in a logistic model could predict the outcome stroke.

In conclusion, this study has revealed a significant high association between gender and stroke; a significant association between high blood pressure and stroke; a positive significant association between the score derived from SES and stroke in both males and females; a negative association between BMI and stroke in men and women as well, however, which was a non-statistically significant association in men but a strong statistically significant association in women, thus supporting the hypothesis that there is indeed a correlation between gender, SES score, blood pressure, BMI, and stroke. These association findings fulfilled one of the several assumptions for a regression model and allowed therefore a hierarchical binary regression to be run. The logistic regression model controlling for age, diabetes, alcohol, and cigarettes consumption revealed age to be a confounding factor for both male and female, however, alcohol consumption was just a confounder for the female gender. Stroke affects both the gender male and female even though there is a distinguishable aspect between the incidence rates and outcomes of these genders. The predictor high blood pressure had a significant influence on stroke, whereas BMI and high SES as compared to middle SES had a non-statistically significant influence on the outcome. Low SES as compared to middle SES had a significant influence on stroke for both males and females. In a nutshell however, this result should be regarded with caution because of several reasons. These are amongst others the study limitations and its overall 70% completed classification table (males: 96%; females: 97.8%). According to Schillmöller (2019), the percentage of correct predicted values in each characteristic (for the case of this study, these are those subjects, who were diagnosed with stroke on one side and those who were not on the other side) has to be above 50% for a classification table to be optimal.

Literature

American Psychological Association. (2018). Socioeconomic status.
Accessed 01. Dec. 2020:
http://www.apa.org/topics/socioeconomic-status/

American Stroke Association (2018). Women Have a Higher Risk of Stroke.
Accessed 25. Jan. 2021:
https://www.stroke.org/en/about-stroke/stroke-risk-factors/women-have-a-higher-risk-of-stroke

Bundeszentrale für politische Bildung. (2020). Bevölkerung nach Altersgruppen und Geschlecht.
Accessed 01. Dec. 2020:
https://www.bpb.de/nachschlagen/zahlen-und-fakten/soziale-situation-in-deutschland/61538/altersgruppen

Appelros, P., Stegmayr, B., & Terent, A. (2009). Sex differences in stroke epidemiology: A systematic review. *Stroke*, *40*(4), 1082–1090. https://doi.org/10.1161/STROKEAHA.108.540781

Avan, A., Digaleh, H., Di Napoli, M., Stranges, S., Behrouz, R., Shojaeianbabaei, G., Amiri, A., Tabrizi, R., Mokhber, N., Spence, J. D., & Azarpazhooh, M. R. (2019). Socioeconomic status and stroke incidence, prevalence, mortality, and worldwide burden: An ecological analysis from the Global Burden of Disease Study 2017. *BMC Medicine*, *17*(1). https://doi.org/10.1186/s12916-019-1397-3

Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research. Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bofah, E. A. tutu, & Hannula, M. S. (2017). Home resources as a measure of socio-economic status in Ghana. *Large-Scale Assessments in Education*, *5*(1), 1–15. https://doi.org/10.1186/s40536-017-0039-5

Broer, M., Bai, Y., & Fonseca, F. (2019). Socioeconomic inequality and educational outcomes: Evidence from twenty years of TIMSS. In *IEA Research for Education* (Vol. 5). https://doi.org/10.1007/978-3-030-11991-1

Busch, M. A., Schienkiewitz, A., Nowossadeck, E., & Gößwald, A. (2013). Prävalenz des Schlaganfalls bei Erwachsenen im Alter von 40 bis 79 Jahren in Deutschland: Ergebnisse der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1). *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *56*(5–6), 656–660. https://doi.org/10.1007/s00103-012-1659-0

Centers for Disease Control and Prevention (2020). Conditions That Increase Risk for Stroke.
Accessed 10. Feb. 2020:
https://www.cdc.gov/stroke/conditions.htm

Centers for disease control and prevention. (2020). Epidemiology.
Accessed 25. Nov. 2020:
https://www.cdc.gov/careerpaths/k12teacherroadmap/epidemiology.html

Centers for disease control and prevention. (2020). Stroke: Stroke Signs and Symptoms.
Accessed 25. Nov. 2020:
https://www.cdc.gov/stroke/signs_symptoms.htm

Columb, M. O., & Atkinson, M. S. (2016). Statistical analysis: sample size and power estimations. *BJA Education*, *16*(5), 159–161. https://doi.org/10.1093/bjaed/mkv034

DeCoster, J. (2004). Data Analysis in SPSS.
Accessed 21. Feb. 2020:
http://www.stat-help.com/notes.html

Deutsches Zentrum für Herz-Kreislaufforschung. (2017). Blood pressure in

Germany is decreasing – yet there is room for much more improvement. Accessed 25. Dec. 2020: https://dzhk.de/en/news/latest-news/article/blood-pressure-in-germany-is-decreasing-yet-there-is-room-for-much-more-improvement/

Engels, T., Baglione, Q., Audibert, M., Viallefont, A., Mourji, F., Faris, M. E. A., Yahyaoui, M., Slassi, I., & Aidi, S. (2014). Socioeconomic status and stroke prevalence in morocco:results from the rabat-casablanca study. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089271

European stroke organization. (2017). Adjustment for confounders in stroke. Accessed 01. Oct. 2020: https://eso-stroke.org/adjustment-confounders-stroke-research/

Fehr, A., Lange, C., Fuchs, J., Neuhauser, H., & Schmitz, R. (2017). Gesundheitsberichterstattung des Bundes gemeinsam getragen von RKI und Destatis. *Journal of Health Monitoring Robert Koch Institut*, *4*(2), 26–30. http://www.gbe-bund.de/pdf/johm_2017_01_gesundheitliche_lage_1.pdf

Feigin, V. L., Krishnamurthi, R. V., Theadom, A. M., Abajobir, A. A., Mishra, S. R., Ahmed, M. B., Abate, K. H., Mengistie, M. A., Wakayo, T., Abd-Allah, F., Abdulle, A. M., Abera, S. F., Mohammed, K. E., Abyu, G. Y., Asgedom, S. W., Atey, T. M., Betsu, B. D., Mezgebe, H. B., Tuem, K. B., … Zaki, M. E. (2017). Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Neurology*, *16*(11), 877–897. https://doi.org/10.1016/S1474-4422(17)30299-5

Fekadu, G., Chelkeba, L., & Kebede, A. (2019). Risk factors, clinical presentations and predictors of stroke among adult patients admitted to stroke unit of Jimma university medical center, south west Ethiopia: prospective observational study. *BMC Neurology*, *19*(1), 1–11. https://doi.org/10.1186/s12883-019-1409-0

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics

Furlan, N. E., Bazan, S. G. Z., Braga, G. P., E Castro, M. C. N., Franco, R. J. da S., Gut, A. L., Bazan, R., & Martin, L. C. (2018). Association between blood pressure and acute phase stroke case fatality rate: A prospective cohort study. *Arquivos de Neuro-Psiquiatria*, *76*(7), 436–443. https://doi.org/10.1590/0004-282x20180059

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. doi:10.1037/a0024338

Hankey, G.J. (2017). Stroke. Lancet. 389(10069): 641- 6

Health Canada. (2000). Health Canada's gender-based analysis policy. Ottawa, 14. Accessed 10. Dec. 2020:
http://www.cwhn.ca/sites/default/files/resources/gba/gender.pdf

Hiraga, A. (2017). Gender Differences and Stroke Outcomes. *Neuroepidemiology*, *48*(1–2), 61–62. https://doi.org/10.1159/000475451

Institute of Health Metrics and Evaluation. (2019).
Accessed 01. Dec. 2020:
http://www.healthdata.org/germany

Institute of Health Metrics and Evaluation. (2019). Global burden of disease compare. Germany.
Accessed 25. Nov. 2020:
http://www.healthdata.org/germany

J Reeves, M., D Bushnell, C., Howard, G., Warner Gargano, J., W Duncan, P., Lynch, G., Khatiwoda, A., & Lisabeth, L. (2011). Sex differences in stroke: epidemiology, clinical presentation,medical care, and outcomes. *National Institutes of Health*, *23*(1), 1–7. https://doi.org/10.1016/S1474-4422(08)70193-5.Sex

Joel C. (2017, Oct. 03). Comparing categorical variables using SPSS [YouTube video].
Accessed 25. Feb. 2020:
https://www.youtube.com/watch?v=nSlzKXe-OGM

Johnston, M. P. (2014). Secondary Data Analysis : A Method of which the Time Has Come. *Qualitatve and Quantative Methods in Libraryes (QQML)*, *3*, 619–626.

Kamtsiuris, P., Lange, M., Hoffmann, R., Schaffrath Rosario, A., Dahm, S., Kuhnert, R., & Kurth, B. M. (2013). Die erste Welle der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1): Stichprobendesign, Response, Gewichtung und Repräsentativität. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *56*(5–6), 620–630. https://doi.org/10.1007/s00103-012-1650-9

Kohler, M., Deutschbein, J., Peschke, D., & Schenk, L. (2014). Stroke occurence in germany - On the comparability of insurance data and registry-based data: Results of a secondary analysis of insurance data (Deutsche BKK). *Fortschritte Der Neurologie Psychiatrie*, *82*(11), 627–633. https://doi.org/10.1055/s-0034-1385231

Kolominsky-Rabas PL, Wiedmann S, Weingärtner M, Liman TG, Endres M, Schwab S, Buchfelder M, Heuschmann PU. (2015). Time trends in incidence of pathological and etiological stroke subtypes during 16 years: the Erlangen Stroke Project. Neuroepidemiology.;44(1):24-9. https://doi: 10.1159/000371353

Kumar, S. (2016). Hypertension and Ischemic Stroke. *Hypertension Journal*, *2*(1), 39–43. https://doi.org/10.5005/jp-journals-10043-0028

Kurth, T., Gaziano, J. M., Rexrode, K. M., Kase, C. S., Cook, N. R., Manson, J. A. E., & Buring, J. E. (2005). Prospective study of body mass index and risk of stroke in apparently healthy women. *Circulation*, *111*(15), 1992–1998. https://doi.org/10.1161/01.CIR.0000161822.83163.B6

Lampert, T., Kroll, L., Müters, S., & Stolzenberg, H. (2013). Messung des soziooökonomischen Status in der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1). *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *56*(5–6), 631–636. https://doi.org/10.1007/s00103-012-1663-4

Lange, C., & Finger, J. D. (2017). Gesundheitsverhalten in Europa - Vergleich ausgewählter Indikatoren für Deutschland und die Europäische Union. *Journal of Health Monitoring FACT SHEET*, *1*(2), 3–20. https://doi.org/10.17886/RKI-GBE-2017-024

Lewington, S., Clarke, R., Qizilbash, N., Peto, R., & Collins, R. (2002). Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet, 360(9349), 1903–1913.

Li, R. C., Xu, W. D., Lei, Y. L., Bao, T., Yang, H. W., Huang, W. X., & Tang, H. R. (2019). The risk of stroke and associated risk factors in a health examination population: A cross-sectional study. *Medicine (United States)*, *98*(40). https://doi.org/10.1097/MD.0000000000017218

Mader, F. M., & Schwenke, R. (2020). Schlaganfall S3-Leitlinie. *Deutsche Gesellschaft Für Allgemeinmedizin Und Familienmedizin (DEGAM), Berlin*, *053*, 185.

Manfred, Te Grotenhuis & Paula, Thijs. (2015). Dummy variables and their interactions in regression analysis: examples from research on body mass index.
Accessed 10. Dec. 2020:
https://www.ru.nl/sociology/mt/bmi/downloads/

Milenović, Ž. (2011). Application of Mann-Whitney U Test in Research of Professional Training of Primary School Teachers. *Metodički Obzori/Methodological Horizons*, *6*(1), 73–79.

https://doi.org/10.32728/mo.06.1.2011.06

National institute of neurological disorders and stroke (2020). Post-Stroke Rehabilitation Fact Sheet. Accessed 10. Jan. 2020: https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Fact-Sheets/Post-Stroke-Rehabilitation-Fact-Sheet

Nimptsch, U., & Mansky, T. (2012). Trends in acute inpatient stroke care in Germany. *Deutsches Arzteblatt International*, *109*(51–52), 885–893. https://doi.org/10.3238/arztebl.2012.0885

O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, Rangarajan S, Islam S, Pais P, McQueen MJ, Mondo C, Damasceno A, Lopez-Jaramillo P, Hankey GJ, Dans AL, Yusoff K, Truelsen T, Diener HC, Sacco RL, Ryglewicz D, Czlonkowska A, Weimar C, Wang X, Yusuf S. (2010). INTERSTROKE investigators. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. Lancet.; 376: 112–123. doi: 10.1016/S0140-6736(10)60834-3

Plass, D., Vos, T., Hornberg, C., Scheidt-Nave, C., Zeeb, H., & Krämer, A. (2014). Entwicklung der Krankheitslast in Deutschland: Ergebnisse, potenziale und grenzen der Global Burden of disease-studie. *Deutsches Arzteblatt International*, *111*(38), 629–638. https://doi.org/10.3238/arztebl.2014.0629

Puthenpurakal, A., & Crussell, J. (2017). Stroke 1: definition, burden, risk factors and diagnosis. *Nursing Times*, *113*(11), 43–47.

Robert Koch-Institut. (2015a). Gesundheitsberichterstattung des Bundes. Gemeinsam getragen von RKI und Destatis, BerGesundheit in Deutschland. *Robert Koch-Institut (Hrsg) (2015) Gesundheit in Deutschland. Gesundheitsberichterstattung Des Bundes. Gemeinsam Getragen von RKI Und Destatis. RKI, Berlin*, 13–76. https://doi.org/10.1016/0168-8510(91)90076-A

Robert Koch-Institut. (2015b). *Studie zur Gesundheit Erwachsener in Deutschland (DEGS1)*. https://doi.org/http://dx.doi.org/10.7797/16-200812-1-1-1

Robert Koch Institut. (2009). DEGS: Studie zur Gesundheit Erwachsener in Deutschland - Projektbeschreibung. *Beiträge Zur Gesundheitsberichtserstattung Des Bundes*. https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsB/degs_projektbeschr.pdf?__blob=publicationFile

Robert Koch Institut. (2013). *Eckdaten DEGS Welle 1 (2008-2011)*. *1*. http://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/Degs/degs_w1/Eckdatenpapier.pdf?__blob=publicationFile

Roth, G. A., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, H. T., Abebe, M., Abebe, Z., Abejie, A. N., Abera, S. F., Abil, O. Z., Abraha, H. N., … Murray, C. J. L. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1736–1788. https://doi.org/10.1016/S0140-6736(18)32203-7

Salvagioni, D. A. J., Melanda, F. N., Mesas, A. E., González, A. D., Gabani, F. L., & De Andrade, S. M. (2017). Physical, psychological and occupational consequences of job burnout: A systematic review of prospective studies. *PLoS ONE*, *12*(10), 1–29. https://doi.org/10.1371/journal.pone.0185781

Schillmöller, Z. (2019). 19-Logistic Regression: Advanced Biostatistics SoSe 2019 [Powerpoint slides].
Accessed 25. Feb. 2021:
https://emil.haw-hamburg.de/mod/folder/view.php?id=321920

Statistics Laerd. (2018) Independent t-test using Spss Statistics.
Accessed 10. Dec. 2020:

https://statistics.laerd.com/spss-tutorials/independent-t-test-using-spss-statistics.php

Stiftung Deutsche Schlaganfall-Hilfe. (n.d.). Mögliche Folgen eines Schlaganfalls. Accessed 25. Nov. 2020: https://www.schlaganfall-hilfe.de/de/verstehen-vermeiden/folgen-eines-schlaganfalls/allgemein#:~:text=L%C3%A4hmungen%20und%20Sprachst%C3%B6rungen%20geh%C3%B6ren%20zu,Gleichgewichtsst%C3%B6rungen%20sowie%20Bewusstseins-%20und%20Wahrnehmungsst%C3%B6rungen.

Substance Abuse and Mental Health Services Administration (n.d.). Risk and Protective Factors
Accessed 30. Jan. 2021: https://www.samhsa.gov/sites/default/files/20190718-samhsa-risk-protective-factors.pdf

Thefeld, W., Stolzenberg, H., & Bellach, B. M. (1999). Bundes-Gesundheitssurvey: Response, Zusammensetzung der Teilnehmer und Non-Responder-Analyse. *Gesundheitswesen*, *61*(SUPPL. 2).

Tran, T. M., & Giang, N. M. (2014). Changes in blood pressure classification, blood pressure goals and pharmacological treatment of essential hypertension in medical guidelines from 2003 to 2013. *IJC Metabolic and Endocrine*, *2*, 1–10. https://doi.org/10.1016/j.ijcme.2014.01.001

Wang, S., Shen, B., Wu, M., Chen, C., & Wang, J. (2019). Effects of socioeconomic status on risk of ischemic stroke: A case-control study in the Guangzhou population. *BMC Public Health*, *19*(1), 1–7. https://doi.org/10.1186/s12889-019-6998-4

Wang C, Liu Y, Yang Q, et al. (2013). Body Mass Index and Risk of Total and Type-Specific Stroke in Chinese Adults: Results from a Longitudinal Study in China. International Journal of Stroke.;8(4):245-250. doi:10.1111/j.1747-4949.2012.00830.x

World Health Organization (WHO). 2020. Body Mass Index.
Accessed 01. Dec. 2020:
https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi

World Health Organization. (2017). Cardiovascular diseases (CVDs).
Accessed 09. Oct. 2020:
https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases (cvds)

World Health Organization. (2020). Determinant of Health. Gender.
Accessed 10. Dec. 2020:
https://www.euro.who.int/en/health-topics/health-determinants/gender/gender#

World Health Organization. (2019). Hypertension
Accessed 25. Dec. 2020:
https://www.who.int/news-room/fact-sheets/detail/hypertension

World Health Organization (1995). Physical Status: The Use and Interpretation of Anthropometry: Report of a World Health Organization (WHO) Expert Committee. Geneva, Switzerland: World Health Organization.

World Health Organization. (n.d.). Risk Factors.
Accessed 30. Jan. 2021:
https://www.who.int/cardiovascular_diseases/en/cvd_atlas_03_risk_factors.pdf

World Stroke Organization (2019). Stroke Riskometer.
Accessed 10. Jan. 2020:
https://www.world-stroke.org/publications-and-resources/resources/stroke-riskometer

Appendices

Appendix I: SPSS Syntax

- **Recoding DEGS1 BMI variable into BMI_group accroding to WHO, 2020**

RECODE BMI (SYSMIS=SYSMIS) (Lowest thru 18.5=1) (18.5 thru 24.9=2) (25.0 thru 29.9=3) (Lowest thru

30=4) INTO BMI_group.

VARIABLE LABELS BMI_group 'BMI group according to WHO'.

EXECUTE.

  - Where:
    1 = Underweight
    2 = Normal weight
    3 = Pre-Obesity or Overweight
    4 = Obesity

- **Descriptive statistics for gender before splitting**

FREQUENCIES VARIABLES=sex

/ORDER=ANALYSIS.

- **Splitting data by gender**

SORT CASES BY sex.

SPLIT FILE LAYERED BY sex.

- **Descriptive statistics**
  - Categorical variables

FREQUENCIES VARIABLES= BMI_group High_blood_pressure SES Stroke

/ORDER=ANALYSIS.

  - Crossing stroke by high blood pressure after splitting by gender

```
EXAMINE VARIABLES=Stroke BY High_blood_pressure
            /PLOT NONE
             /STATISTICS DESCRIPTIVES
             /CINTERVAL 95
             /MISSING LISTWISE
             /NOTOTAL.
```

- Numerical variables

```
FREQUENCIES VARIABLES=SES_score BMI
           /FORMAT=NOTABLE
            /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM SEMEAN MEAN
            /HISTOGRAM
             /ORDER=ANALYSIS.
```

- Description of SES_score and BMI after splitting

```
FREQUENCIES VARIABLES=SES_score BMI
           /FORMAT=NOTABLE
            /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM SEMEAN MEAN
           /ORDER=ANALYSIS.
```

- **Kolmogorov-Smirnov test for normal distribution**

```
EXAMINE VARIABLES=SES_score BMI
             /PLOT BOXPLOT HISTOGRAM NPPLOT
             /COMPARE GROUPS
             /STATISTICS DESCRIPTIVES
             /CINTERVAL 95
             /MISSING LISTWISE
             /NOTOTAL.
```

- **Assessing differences**

  – Crosstabulation for sex without splitting gender

```
CROSSTABS
  /TABLES=Stroke BY sex
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT ROW COLUMN TOTAL
  /COUNT ROUND CELL.
```

  – Crosstabulation for High_blood_pressure

```
CROSSTABS
  /TABLES=Stroke BY High_blood_pressure
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT ROW COLUMN TOTAL
  /COUNT ROUND CELL.
```

  – Mann Whitney U-Test

```
NPAR TESTS
  /M-W= SES_score BMI BY Stroke(1 2)
  /MISSING LISTWISE.
```

- **Correlations**

  – Point-Biserial correlation with pearson

```
CORRELATIONS
  /VARIABLES=Stroke SES_score BMI
  /PRINT=TWOTAIL NOSIG FULL
  /STATISTICS DESCRIPTIVES
  /MISSING=LISTWISE.
```

- Chi-Square for Stroke and sex without split

```
CROSSTABS
  /TABLES=Stroke BY sex
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.
```

- Chi-Square for Stroke and High_blood_pressure after case sorted by gender

```
CROSSTABS
  /TABLES=Stroke BY High_blood_pressure
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.
```

- **Coding SES to dummy variables**

```
RECODE SES (1=1) (SYSMIS=SYSMIS) (ELSE=0) INTO Low_SES.
VARIABLE LABELS  Low_SES 'Low SES'.
EXECUTE.

RECODE SES (2=1) (SYSMIS=SYSMIS) (ELSE=0) INTO Middle_SES.
VARIABLE LABELS  Middle_SES 'Middle SES'.
EXECUTE.

RECODE SES (3=1) (SYSMIS=SYSMIS) (ELSE=0) INTO High_SES.
VARIABLE LABELS  High_SES 'High SES'.
EXECUTE.
```

- **Computing means variables by aggregate**

```
AGGREGATE
    /OUTFILE=* MODE=ADDVARIABLES
    /BREAK=
    /SES_score_mean=MEAN(SES_score)
    /BMI_mean=MEAN(BMI).
```

- Mean centering variables

```
COMPUTE BMI_mean_centered=BMI - BMI_mean.
EXECUTE.

COMPUTE SES_score_mean_centered=SES_score - SES_score_mean.
EXECUTE.
```

- Computing interaction between BMI_mean_centered and SES_score_mean_centered

```
COMPUTE Mod_BMI_SES=BMI_mean_centered * SES_score_mean_centered.
EXECUTE.
```

- Computing interaction between BMI_mean_centered and low and high_SES_dummy

```
COMPUTE Mod_BMI_Low_SES_Dummy=BMI_mean_centered * Low_SES.
VARIABLE LABELS  Mod_BMI_Low_SES_Dummy 'Mod_BMI Low SES Dummy'.
EXECUTE.

COMPUTE Mod_BMI_High_SES_Dummy=BMI_mean_centered * High_SES.
VARIABLE LABELS  Mod_BMI_High_SES_Dummy 'Mod BMI High SES Dummy'.
EXECUTE.
```

- **Hierarchical multiple binary logistic regression**

```
COMPUTE Age=2007 - gebj.
EXECUTE.
```

– Regression with gender as predictor before splitting by sex

The first model had to include the confounders; the predictors incrementally added to detect any mediation effects

```
LOGISTIC REGRESSION VARIABLES Stroke
        /METHOD=ENTER Age Alcohol Smoking Diabetes
        /METHOD=ENTER sex High_blood_pressure Low_SES High_SES BMI_mean_centered
        /METHOD=ENTER Mod_BMI_SES
        /CONTRAST (Alcohol)=Indicator
        /CONTRAST (Diabetes)=Indicator
        /CONTRAST (Smoking)=Indicator
        /CONTRAST (sex)=Indicator
        /CONTRAST (High_blood_pressure)=Indicator
        /SAVE=COOK
        /CLASSPLOT
        /PRINT=GOODFIT CI(95)
        /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(50) CUT(0.5).
```

– Regression without sex as predictor, after splitting by sex,

WITH middle SES as reference of the dummies

```
LOGISTIC REGRESSION VARIABLES Stroke
        /METHOD=ENTER Age Alcohol Smoking Diabetes
        /METHOD=ENTER High_blood_pressure Low_SES High_SES BMI_mean_centered
        /METHOD=ENTER Mod_BMI_SES
        /CONTRAST (Alcohol)=Indicator
        /CONTRAST (Diabetes)=Indicator
        /CONTRAST (Smoking)=Indicator
        /CONTRAST (High_blood_pressure)=Indicator
        /SAVE=COOK
```

```
    /CLASSPLOT
    /PRINT=GOODFIT CI(95)
    /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(50) CUT(0.5).
```

***Cut off set at 0.8*

```
LOGISTIC REGRESSION VARIABLES Stroke
    /METHOD=ENTER Age Alcohol Smoking Diabetes
    /METHOD=ENTER High_blood_pressure Low_SES High_SES BMI_mean_centered
    /METHOD=ENTER Mod_BMI_SES
    /CONTRAST (Alcohol)=Indicator
    /CONTRAST (Diabetes)=Indicator
    /CONTRAST (Smoking)=Indicator
    /CONTRAST (High_blood_pressure)=Indicator
    /SAVE=COOK
    /CLASSPLOT
    /PRINT=GOODFIT CI(95)
    /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(50) CUT(0.8).
```

***Cut off set at 0.9*

```
LOGISTIC REGRESSION VARIABLES Stroke
    /METHOD=ENTER Age Alcohol Smoking Diabetes
    /METHOD=ENTER High_blood_pressure Low_SES High_SES BMI_mean_centered
    /METHOD=ENTER Mod_BMI_SES
    /CONTRAST (Alcohol)=Indicator
    /CONTRAST (Diabetes)=Indicator
    /CONTRAST (Smoking)=Indicator
    /CONTRAST (High_blood_pressure)=Indicator
    /SAVE=COOK
    /CLASSPLOT
    /PRINT=GOODFIT CI(95)
    /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(50) CUT(0.9).
```

- **VIF: Linear regression for multicollinearity assessment for males and females**

```
REGRESSION
          /MISSING LISTWISE
          /STATISTICS COLLIN TOL
           /CRITERIA=PIN(.05) POUT(.10)
           /NOORIGIN
           /DEPENDENT Stroke
           /METHOD=ENTER BMI SES_score.
```

Appendix II: Complete model estimates for regression on stroke including the variable gender

Models	Variables	B	S.E.	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
I	Age	-0.066	0.010	1	<0.001	,936	,917	,956
	Alcohol consumption	0.238	0.206	1	0.249	1,268	,847	1,900
	Cigarette consumption	-0.113	0.285	1	0.691	,893	,511	1,560
	Diabetes	-0.468	0.209	1	0.025	,626	,416	,943
II	Gender	-0.646	0.182		<0.001	,524	,367	,748
	High blood pressure	-1.253	0.238	1	<0,001	,286	,179	,455
	Low SES	-0.631	0.209	1	0.003	,532	,353	,802
	High SES	0.479	0.254	1	0.059	1,614	,981	2,654
	BMI	0.008	0.020	1	0.708	1,008	,968	1,049
III	Mod BMI SES	-0.001	0.005	1	0.812	,999	,989	1,009

Appendix III: Overview of the variables, level of measurements and reference group used in the statistics and their corresponding variables in the dataset DEGS1

Variable names in this study	Variable names in DEGS1	Type of data, as appearing in DEGS1	Modified data level for current study	Reference group
Gender	Sex	Nominal dichotomous	-	
SES	SDses	Ordinal (3 Levels)	Dummy coded	Low-SES
SES-Score	SDses_score	Continuous metric	-	-
Blood pressure	KHhyp	Nominal dichotomous	-	
BMI	USbmi_st	Continuous	Continuous	
BMI group	-	-	Categorical Ordinal	-
Stroke	KHsa	Nominal dichotomous	-	
Confounder				
Age	Age	-	-	-
Alcohol consumption	AKriskoB	Nominal dichotomous	-	-
Smoking	USbdrauch	Nominal dichotomous	-	-
Diabetes	KHdiab	Nominal dichotomous	-	-

Appendix IV: Supplement tables of Binary regression analyses without gender splitting

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	4841	60,6
	Missing Cases	3146	39,4
	Total	7987	100,0
Unselected Cases		0	,0
Total		7987	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Yes	0
No	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
High blood pressure (Medically diagnosed)	Yes	2331	1,000
	No	2510	,000
Cigarette consumption	Yes	736	1,000
	No	4105	,000
Diabetes (medically diagnosed)	Yes	497	1,000
	No	4344	,000
Gender	Male	2306	1,000
	Female	2535	,000
Alcohol consumption	Yes	1490	1,000
	No	3351	,000

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted: Stroke (Medically diagnosed) Yes	Predicted: Stroke (Medically diagnosed) No	Predicted: Percentage Correct
Step 0	Stroke (Medically diagnosed)	Yes	0	147	,0
		No	0	4694	100,0
	Overall Percentage				97,0

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	3,464	,084	1709,947	1	,000	31,932

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Age	97,019	1	,000
		Alcohol consumption(1)	3,456	1	,063
		Cigarette consumption (1)	2,194	1	,139
		Diabetes (medically diagnosed)(1)	43,531	1	,000
	Overall Statistics		120,263	4	,000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	117,314	4	,000
	Block	117,314	4	,000
	Model	117,314	4	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1199,543[a]	,024	,101

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	11,540	8	,173

Contingency Table for Hosmer and Lemeshow Test

		Stroke (Medically diagnosed) = Yes		Stroke (Medically diagnosed) = No		
		Observed	Expected	Observed	Expected	Total
Step 1	1	43	49,368	464	457,632	507
	2	37	27,336	447	456,664	484
	3	21	20,882	470	470,118	491
	4	17	15,934	495	496,066	512
	5	10	11,148	503	501,852	513
	6	5	7,763	489	486,237	494
	7	4	5,460	472	470,540	476
	8	8	4,266	488	491,734	496
	9	2	3,153	495	493,847	497
	10	0	1,691	371	369,309	371

Classification Table[a]

			Predicted		
			Stroke (Medically diagnosed)		
Observed			Yes	No	Percentage Correct
Step 1	Stroke (Medically diagnosed)	Yes	0	147	,0
		No	0	4694	100,0
	Overall Percentage				97,0

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Age	-,083	,010	67,451	1	,000	,920	,902	,939
	Alcohol consumption(1)	,164	,200	,673	1	,412	1,179	,796	1,746
	Cigarette consumption (1)	-,184	,279	,433	1	,510	,832	,481	1,439
	Diabetes (medically diagnosed)(1)	-,729	,200	13,234	1	,000	,483	,326	,715
	Constant	8,598	,662	168,452	1	,000	5422,423		

a. Variable(s) entered on step 1: Age, Alcohol consumption, Cigarette consumption , Diabetes (medically diagnosed).

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	60,001	5	,000
	Block	60,001	5	,000
	Model	177,315	9	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1139,542[a]	,036	,151

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6,672	8	,572

Contingency Table for Hosmer and Lemeshow Test

		Stroke (Medically diagnosed) = Yes		Stroke (Medically diagnosed) = No		
		Observed	Expected	Observed	Expected	Total
Step 1	1	55	58,596	429	425,404	484
	2	36	30,118	448	453,882	484
	3	22	19,958	462	464,042	484
	4	13	13,333	471	470,667	484
	5	10	9,040	474	474,960	484
	6	4	6,149	480	477,851	484
	7	1	4,179	483	479,821	484
	8	4	2,746	480	481,254	484
	9	2	1,814	482	482,186	484
	10	0	1,068	485	483,932	485

Classification Table[a]

			Predicted		
			Stroke (Medically diagnosed)		
	Observed		Yes	No	Percentage Correct
Step 1	Stroke (Medically diagnosed)	Yes	0	147	,0
		No	0	4694	100,0
	Overall Percentage				97,0

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Age	-,066	,010	39,520	1	,000	,936	,917	,956
	Alcohol consumption(1)	,237	,206	1,318	1	,251	1,267	,846	1,898
	Cigarette consumption (1)	-,113	,285	,157	1	,692	,893	,511	1,560
	Diabetes (medically diagnosed)(1)	-,469	,209	5,023	1	,025	,626	,415	,943
	Gender(1)	-,649	,182	12,775	1	,000	,523	,366	,746
	High blood pressure (Medically diagnosed)(1)	-1,255	,238	27,885	1	,000	,285	,179	,454
	Low SES	-,619	,204	9,243	1	,002	,539	,361	,803
	High SES	,473	,253	3,506	1	,061	1,605	,978	2,632
	BMI mean centered	,009	,019	,246	1	,620	1,010	,972	1,048
	Constant	8,695	,686	160,756	1	,000	5971,711		

a. Variable(s) entered on step 1: Gender, High blood pressure (Medically diagnosed), Low SES, High SES, BMI mean centered.

Block 3: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	,056	1	,812
	Block	,056	1	,812
	Model	177,371	10	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1139,486[a]	,036	,151

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5,038	8	,754

Contingency Table for Hosmer and Lemeshow Test

		Stroke (Medically diagnosed) = Yes		Stroke (Medically diagnosed) = No		
		Observed	Expected	Observed	Expected	Total
Step 1	1	55	58,527	429	425,473	484
	2	37	30,155	447	453,845	484
	3	20	19,985	464	464,015	484
	4	14	13,366	470	470,634	484
	5	10	9,050	474	474,950	484
	6	4	6,149	480	477,851	484
	7	2	4,175	482	479,825	484
	8	3	2,739	481	481,261	484
	9	2	1,801	482	482,199	484
	10	0	1,054	485	483,946	485

Classification Table[a]

Observed			Predicted: Stroke (Medically diagnosed) Yes	Predicted: Stroke (Medically diagnosed) No	Predicted: Percentage Correct
Step 1	Stroke (Medically diagnosed)	Yes	0	147	,0
		No	0	4694	100,0
	Overall Percentage				97,0

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Age	-,066	,010	39,539	1	,000	,936	,917	,956
	Alcohol consumption(1)	,238	,206	1,328	1	,249	1,268	,847	1,900
	Cigarette consumption (1)	-,113	,285	,158	1	,691	,893	,511	1,560
	Diabetes (medically diagnosed)(1)	-,468	,209	5,021	1	,025	,626	,416	,943
	Gender(1)	-,646	,182	12,651	1	,000	,524	,367	,748
	High blood pressure (Medically diagnosed)(1)	-1,253	,238	27,816	1	,000	,286	,179	,455
	Low SES	-,631	,209	9,083	1	,003	,532	,353	,802
	High SES	,479	,254	3,556	1	,059	1,614	,981	2,654
	BMI mean centered	,008	,020	,140	1	,708	1,008	,968	1,049
	Mod BMI SES	-,001	,005	,057	1	,812	,999	,989	1,009
	Constant	8,695	,686	160,680	1	,000	5972,934		

a. Variable(s) entered on step 1: Mod BMI SES.

Appendix V: Supplement tables of Binary regression analyses for males and females

Logistic Regression

Case Processing Summary

Gender	Unweighted Cases[a]		N	Percent
Male	Selected Cases	Included in Analysis	2306	60,9
		Missing Cases	1483	39,1
		Total	3789	100,0
	Unselected Cases		0	,0
	Total		3789	100,0
Female	Selected Cases	Included in Analysis	2535	60,4
		Missing Cases	1663	39,6
		Total	4198	100,0
	Unselected Cases		0	,0
	Total		4198	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Gender	Original Value	Internal Value
Male	Yes	0
	No	1
Female	Yes	0
	No	1

Categorical Variables Codings

Gender			Frequency	Parameter coding (1)
Male	High blood pressure (Medically diagnosed)	Yes	1179	1,000
		No	1127	,000
	Cigarette consumption	Yes	366	1,000
		No	1940	,000
	Diabetes (medically diagnosed)	Yes	274	1,000
		No	2032	,000
	Alcohol consumption	Yes	873	1,000
		No	1433	,000

Female	High blood pressure (Medically diagnosed)	Yes	1152	1,000
		No	1383	,000
	Cigarette consumption	Yes	370	1,000
		No	2165	,000
	Diabetes (medically diagnosed)	Yes	223	1,000
		No	2312	,000
	Alcohol consumption	Yes	617	1,000
		No	1918	,000

Block 0: Beginning Block

Classification Table[a,b]

				Predicted		
				Stroke (Medically diagnosed)		
Gender		Observed		Yes	No	Percentage Correct
Male	Step 0	Stroke (Medically diagnosed)	Yes	0	92	,0
			No	0	2214	100,0
		Overall Percentage				96,0
Female	Step 0	Stroke (Medically diagnosed)	Yes	0	55	,0
			No	0	2480	100,0
		Overall Percentage				97,8

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

Gender			B	S.E.	Wald	df	Sig.	Exp(B)
Male	Step 0[a]	Constant	3,181	,106	893,655	1	,000	24,065
Female	Step 0[a]	Constant	3,809	,136	780,523	1	,000	45,091

a. Variable(s) entered on step 1: Mod BMI SES.

Variables not in the Equation

Gender				Score	df	Sig.
Male	Step 0	Variables	Age	54,361	1	,000
			Alcohol consumption(1)	,385	1	,535
			Cigarette consumption (1)	,218	1	,641
			Diabetes (medically diagnosed)(1)	15,749	1	,000

		Overall Statistics		60,901	4	,000
Female	Step 0	Variables	Age	39,662	1	,000
			Alcohol consumption(1)	10,888	1	,001
			Cigarette consumption (1)	3,769	1	,052
			Diabetes (medically diagnosed)(1)	28,860	1	,000
		Overall Statistics		65,812	4	,000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

Gender			Chi-square	df	Sig.
Male	Step 1	Step	63,623	4	,000
		Block	63,623	4	,000
		Model	63,623	4	,000
Female	Step 1	Step	61,414	4	,000
		Block	61,414	4	,000
		Model	61,414	4	,000

Model Summary

Gender	Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Male	1	709,409[a]	,027	,096
Female	1	468,752[b]	,024	,127

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001 for split file Gender = Male.

b. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001 for split file Gender = Female.

Hosmer and Lemeshow Test

Gender	Step	Chi-square	df	Sig.
Male	1	9,961	8	,268
Female	1	15,645	8	,048

Contingency Table for Hosmer and Lemeshow Test

Gender			Stroke (Medically diagnosed) = Yes Observed	Stroke (Medically diagnosed) = Yes Expected	Stroke (Medically diagnosed) = No Observed	Stroke (Medically diagnosed) = No Expected	Total
Male	Step 1	1	20	26,317	205	198,683	225
		2	17	17,172	206	205,828	223
		3	22	13,084	197	205,916	219
		4	10	10,967	229	228,033	239
		5	8	7,704	221	221,296	229
		6	6	5,700	224	224,300	230
		7	4	4,074	218	217,926	222
		8	3	3,140	237	236,860	240
		9	2	2,218	228	227,782	230
		10	0	1,623	249	247,377	249
Female	Step 1	1	20	21,187	228	226,813	248
		2	15	10,217	218	222,783	233
		3	7	7,836	247	246,164	254
		4	2	5,063	250	246,937	252
		5	2	3,479	256	254,521	258
		6	2	2,496	254	253,504	256
		7	6	1,927	256	260,073	262
		8	0	1,412	257	255,588	257
		9	1	,965	258	258,035	259
		10	0	,416	256	255,584	256

Classification Table[a]

Gender		Observed		Predicted: Stroke (Medically diagnosed) Yes	Predicted: Stroke (Medically diagnosed) No	Predicted: Percentage Correct
Male	Step 1	Stroke (Medically diagnosed)	Yes	0	92	,0
			No	0	2214	100,0
		Overall Percentage				96,0
Female	Step 1	Stroke (Medically diagnosed)	Yes	0	55	,0
			No	0	2480	100,0
		Overall Percentage				97,8

a. The cut value is ,500

Variables in the Equation

Gender			B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
									Lower	Upper
Male	Step 1[a]	Age	-,084	,013	42,498	1	,000	,919	,896	,943
		Alcohol consumption(1)	-,008	,229	,001	1	,973	,992	,634	1,554
		Cigarette consumption (1)	-,451	,321	1,967	1	,161	,637	,339	1,196
		Diabetes (medically diagnosed)(1)	-,491	,259	3,595	1	,058	,612	,369	1,017
		Constant	8,453	,851	98,699	1	,000	4687,428		
Female	Step 1[a]	Age	-,075	,016	21,842	1	,000	,927	,898	,957
		Alcohol consumption(1)	1,410	,601	5,507	1	,019	4,097	1,262	13,302
		Cigarette consumption (1)	,544	,612	,790	1	,374	1,723	,519	5,715
		Diabetes (medically diagnosed)(1)	-,978	,319	9,426	1	,002	,376	,201	,702
		Constant	8,280	1,042	63,163	1	,000	3943,471		

a. Variable(s) entered on step 1: Age, Alcohol consumption, Cigarette consumption , Diabetes (medically diagnosed).

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

Gender			Chi-square	df	Sig.
Male	Step 1	Step	25,732	4	,000
		Block	25,732	4	,000
		Model	89,354	8	,000
Female	Step 1	Step	24,162	4	,000
		Block	24,162	4	,000
		Model	85,575	8	,000

Model Summary

Gender	Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Male	1	683,678[a]	,038	,133
Female	1	444,590[b]	,033	,176

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001 for split file Gender = Male.

b. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001 for split file Gender = Female.

Hosmer and Lemeshow Test

Gender	Step	Chi-square	df	Sig.
Male	1	7,551	8	,479
Female	1	7,204	8	,515

Contingency Table for Hosmer and Lemeshow Test

			Stroke (Medically diagnosed) = Yes		Stroke (Medically diagnosed) = No		
Gender			Observed	Expected	Observed	Expected	Total
Male	Step 1	1	29	32,055	202	198,945	231
		2	20	18,690	211	212,310	231
		3	18	13,190	213	217,810	231
		4	7	9,191	224	221,809	231
		5	6	6,359	225	224,641	231
		6	7	4,562	224	226,438	231
		7	3	3,297	228	227,703	231
		8	0	2,284	231	228,716	231
		9	2	1,510	229	229,490	231
		10	0	,862	227	226,138	227
Female	Step 1	1	25	25,049	229	228,951	254
		2	15	12,482	239	241,518	254
		3	4	7,112	250	246,888	254
		4	6	3,960	248	250,040	254
		5	0	2,425	254	251,575	254
		6	2	1,529	252	252,471	254
		7	2	1,050	252	252,950	254
		8	1	,741	253	253,259	254
		9	0	,453	254	253,547	254
		10	0	,199	249	248,801	249

Classification Table[a]

				Predicted		
				Stroke (Medically diagnosed)		
Gender		Observed		Yes	No	Percentage Correct
Male	Step 1	Stroke (Medically diagnosed)	Yes	0	92	,0
			No	0	2214	100,0
		Overall Percentage				96,0
Female	Step 1	Stroke (Medically diagnosed)	Yes	0	55	,0
			No	0	2480	100,0
		Overall Percentage				97,8

a. The cut value is ,500

Variables in the Equation

Gender			B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Male	Step 1[a]	Age	-,072	,013	29,297	1	,000	,931	,907	,955
		Alcohol consumption(1)	-,015	,232	,004	1	,950	,985	,626	1,552
		Cigarette consumption (1)	-,369	,328	1,265	1	,261	,691	,363	1,316
		Diabetes (medically diagnosed)(1)	-,365	,270	1,833	1	,176	,694	,409	1,177
		High blood pressure (Medically diagnosed)(1)	-1,121	,287	15,275	1	,000	,326	,186	,572
		Low SES	-,635	,272	5,448	1	,020	,530	,311	,903
		High SES	,339	,284	1,424	1	,233	1,404	,804	2,452
		BMI mean centered	,026	,028	,883	1	,347	1,026	,972	1,084
		Constant	8,421	,873	92,971	1	,000	4540,227		
Female	Step 1[a]	Age	-,051	,017	9,019	1	,003	,950	,919	,982
		Alcohol consumption(1)	1,229	,605	4,129	1	,042	3,419	1,045	11,190
		Cigarette consumption (1)	,596	,618	,930	1	,335	1,815	,540	6,101
		Diabetes (medically diagnosed)(1)	-,667	,332	4,034	1	,045	,513	,268	,984
		High blood pressure (Medically diagnosed)(1)	-1,483	,431	11,856	1	,001	,227	,098	,528
		Low SES	-,547	,308	3,148	1	,076	,579	,316	1,059
		High SES	,949	,613	2,401	1	,121	2,584	,778	8,587
		BMI mean centered	-,001	,027	,002	1	,964	,999	,948	1,052
		Constant	7,826	1,086	51,908	1	,000	2505,782		

a. Variable(s) entered on step 1: High blood pressure (Medically diagnosed), Low SES, High SES, BMI mean centered.

Block 3: Method = Enter

Omnibus Tests of Model Coefficients

Gender			Chi-square	df	Sig.
Male	Step 1	Step	,490	1	,484
		Block	,490	1	,484
		Model	89,845	9	,000
Female	Step 1	Step	1,477	1	,224
		Block	1,477	1	,224
		Model	87,053	9	,000

Model Summary

Gender	Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Male	1	683,188[a]	,038	,134
Female	1	443,113[b]	,034	,179

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001 for split file Gender = Male.

b. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001 for split file Gender = Female.

Hosmer and Lemeshow Test

Gender	Step	Chi-square	df	Sig.
Male	1	10,978	8	,203
Female	1	6,188	8	,626

Contingency Table for Hosmer and Lemeshow Test

Gender			Stroke (Medically diagnosed) = Yes		Stroke (Medically diagnosed) = No		
			Observed	Expected	Observed	Expected	Total
Male	Step 1	1	29	32,240	202	198,760	231
		2	21	18,708	210	212,292	231
		3	17	13,139	214	217,861	231
		4	7	9,123	224	221,877	231
		5	5	6,323	226	224,677	231
		6	9	4,543	222	226,457	231
		7	2	3,273	229	227,727	231
		8	0	2,279	231	228,721	231
		9	2	1,509	229	229,491	231
		10	0	,865	227	226,135	227
Female	Step 1	1	24	25,071	230	228,929	254
		2	16	12,650	238	241,350	254
		3	4	7,164	250	246,836	254
		4	6	3,941	248	250,059	254
		5	1	2,400	253	251,600	254
		6	1	1,505	253	252,495	254
		7	2	1,004	252	252,996	254
		8	1	,704	253	253,296	254
		9	0	,401	254	253,599	254
		10	0	,160	249	248,840	249

Classification Table[a]

Gender		Observed		Predicted: Stroke (Medically diagnosed) Yes	Predicted: Stroke (Medically diagnosed) No	Predicted: Percentage Correct
Male	Step 1	Stroke (Medically diagnosed)	Yes	0	92	,0
			No	0	2214	100,0
		Overall Percentage				96,0
Female	Step 1	Stroke (Medically diagnosed)	Yes	0	55	,0
			No	0	2480	100,0
		Overall Percentage				97,8

a. The cut value is ,500

Variables in the Equation

Gender			B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Male	Step 1[a]	Age	-,072	,013	29,290	1	,000	,931	,907	,955
		Alcohol consumption(1)	-,018	,232	,006	1	,939	,982	,624	1,547
		Cigarette consumption (1)	-,373	,328	1,295	1	,255	,689	,362	1,309
		Diabetes (medically diagnosed)(1)	-,361	,270	1,795	1	,180	,697	,411	1,182
		High blood pressure (Medically diagnosed)(1)	-1,124	,287	15,343	1	,000	,325	,185	,570
		Low SES	-,600	,277	4,694	1	,030	,549	,319	,944
		High SES	,324	,286	1,282	1	,257	1,383	,789	2,422
		BMI mean centered	,032	,029	1,203	1	,273	1,033	,975	1,094
		Mod BMI SES	,005	,007	,486	1	,486	1,005	,991	1,019
		Constant	8,426	,874	93,039	1	,000	4562,744		
Female	Step 1[a]	Age	-,052	,017	9,268	1	,002	,949	,918	,982
		Alcohol consumption(1)	1,231	,605	4,140	1	,042	3,426	1,046	11,218
		Cigarette consumption (1)	,570	,620	,846	1	,358	1,768	,525	5,957
		Diabetes (medically diagnosed)(1)	-,660	,331	3,978	1	,046	,517	,270	,989
		High blood pressure (Medically diagnosed)(1)	-1,464	,431	11,535	1	,001	,231	,099	,538
		Low SES	-,675	,322	4,403	1	,036	,509	,271	,956
		High SES	,997	,617	2,611	1	,106	2,710	,809	9,082
		BMI mean centered	-,019	,029	,433	1	,511	,981	,927	1,039
		Mod BMI SES	-,010	,008	1,527	1	,217	,990	,974	1,006
		Constant	7,893	1,096	51,894	1	,000	2678,281		

a. Variable(s) entered on step 1: Mod BMI SES.

Appendix VI: Linear regression for multicollinearity assessment for males and females

Regression

Variables Entered/Removed[a]

Gender	Model	Variables Entered	Variables Removed	Method
Male	1	SES-Score, Body-Mass-Index [kg/m2][b]	.	Enter
Female	1	SES-Score, Body-Mass-Index [kg/m2][b]	.	Enter

a. Dependent Variable: Stroke (Medically diagnosed)

b. All requested variables entered.

Coefficients[a]

			Collinearity Statistics	
Gender	Model		Tolerance	VIF
Male	1	Body-Mass-Index [kg/m2]	,988	1,012
		SES-Score	,988	1,012
Female	1	Body-Mass-Index [kg/m2]	,941	1,063
		SES-Score	,941	1,063

a. Dependent Variable: Stroke (Medically diagnosed)

Collinearity Diagnostics[a]

					Variance Proportions		
Gender	Model	Dimension	Eigenvalue	Condition Index	(Constant)	Body-Mass-Index [kg/m2]	SES-Score
Male	1	1	2,915	1,000	,00	,00	,01
		2	,075	6,255	,02	,08	,85
		3	,010	17,026	,98	,92	,14

Female	1	1	2,904	1,000	,00	,00	,01
		2	,082	5,944	,01	,15	,64
		3	,013	14,671	,99	,84	,35

a. Dependent Variable: Stroke (Medically diagnosed)

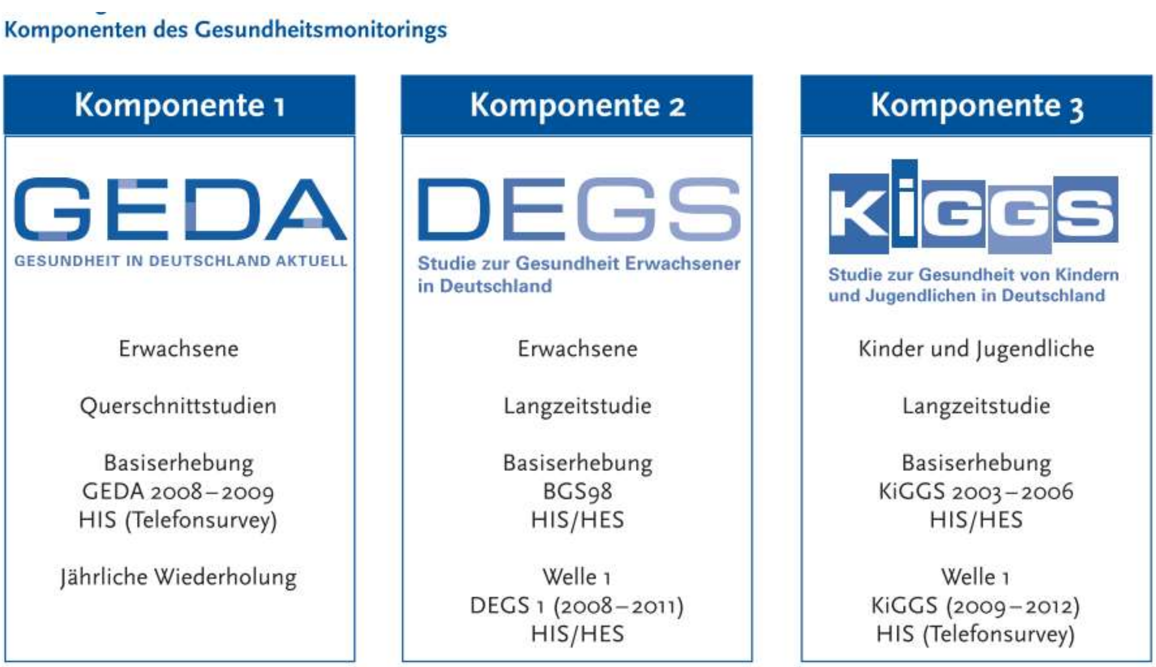
Appendix VII: RKI Nutzung von Scientific UseFiles

Nutzung von Scientific Use Files

Folgende Regelungen und Verpflichtungen sind durch den Antragsteller/die Antragstellerin einzuhalten:

- Verwendung des überlassenen Datensatzes nur für die beantragten Zwecke, insbesondere Ausschluss einer kommerziellen Nutzung des Datensatzes.
- Ausschluss der Weitergabe dieser Daten an Dritte und Hinweis, dass in diesem Fall ein gesonderter Antrag erforderlich ist.
- Einhaltung der wesentlichen Regeln des anerkannten Standes von Wissenschaft und Technik („Gute Epidemiologische Praxis"[1] bzw. „Gute Praxis Sekundärdatenanalyse"[2], s. unten), soweit auf den Datensatz anwendbar.
- Verpflichtung zur Einhaltung allgemeiner Datenschutzanforderungen, insbesondere kein Versuch der Deanonymisierung der Daten, der Verlinkung von Einzelfalldaten mit anderen Datenquellen sowie Publikation von Einzelfalldaten.
- Verpflichtung zur Nennung der Datenquelle bei Publikation.
  (siehe Informationsblatt zum jeweiligen SUF)
- Verpflichtung zur Mitteilung an das RKI-Forschungsdatenzentrum (fdz@rki.de) über erfolgte Publikationen mit der Bitte um Zusendung eines elektronischen Belegexemplars jeder Veröffentlichung zu Dokumentationszwecken.
- Bitte um Hinweise auf evtl. Fehler in den Daten oder in der Dokumentation.
- Ich habe die Datenschutzerklärung des RKI zur Kenntnis genommen und bin damit einverstanden, dass die von mir angegebenen Daten elektronisch erhoben und gespeichert werden. Meine Daten werden dabei nur zur Bearbeitung und Beantwortung meiner Datenanfrage benutzt und nicht an Dritte weitergegeben.

Appendix VIII: Different components of the RKI health monitoring program



Appendix IX: Key data DEGS first wave

ROBERT KOCH INSTITUT

Eckdaten DEGS Welle 1 (2008-2011)

Studie zur Gesundheit Erwachsener in Deutschland

Verantwortliche Durchführung

- Robert Koch-Institut, Abteilung für Epidemiologie und Gesundheitsberichterstattung, Postfach 650261, 13302 Berlin.

Auftraggeber

- Bundesministerium für Gesundheit

Ziele der Studie

- Die „Studie zur Gesundheit Erwachsener in Deutschland" (DEGS) ist eine bundesweite Quer- und Längsschnitterhebung, die vom Robert Koch-Institut (RKI) im Rahmen des Gesundheitsmonitorings durchgeführt wird. Ziel des Gesundheitsmonitorings ist die kontinuierliche Bereitstellung bundesweit repräsentativer Daten zur Gesundheit der Bevölkerung, die Ermittlung zeitlicher Trends in der Entwicklung der gesundheitlichen Lage und die Sammlung von Erkenntnissen über die gesundheitliche Entwicklung im Lebensverlauf.
- Mit der ersten Welle von DEGS (DEGS1) hat das RKI erstmals seit dem Bundes-Gesundheitssurvey 1998 (BGS98) bundesweit gültige Untersuchungs- und Befragungsdaten zur Gesundheit der erwachsenen Bevölkerung erhoben.
- Die Daten ermöglichen Aussagen über die aktuelle gesundheitliche Situation der Erwachsenen in Deutschland und – im Vergleich mit den Ergebnissen des BGS98 – Aussagen über Trends in der Entwicklung der gesundheitlichen Lage. Die erneute Einbeziehung der Studienteilnehmer des BGS98 ermöglicht Erkenntnisse über ursächliche Zusammenhänge von Gesundheitsverhalten, Unterstützungs- und Risikopotenzialen und dem aktuellen Gesundheitszustand im Lebensverlauf.
- Die gewonnenen Daten bilden eine wichtige Grundlage für die Gesundheitsberichterstattung auf Bundesebene und sind grundlegend für die bedarfsgerechte gesundheitspolitische Planung und für Prioritätensetzungen zur Verbesserung der gesundheitlichen Situation der Bevölkerung.
- Die Daten dienen der epidemiologischen Forschung, bspw. zur Berechnung von Risikomodellen, zur Identifizierung von gesund erhaltenden und krank machenden Faktoren sowie für komplexe Zusammenhangsanalysen.

Laufzeit DEGS1

- IV. Quartal 2008 bis IV. Quartal 2011 (Untersuchungszeitraum)

Themen und thematische Schwerpunkte der Studie

- Gesundheitsstatus (Krankheiten, Verletzungen/Vergiftungen, psychische Gesundheit und physiologische Risikofaktoren)
- Subjektive Gesundheit und gesundheitsbezogene Lebensqualität
- Inanspruchnahme von Leistungen des Gesundheitssystems (einschließlich Leistungen zur Primär- und Sekundärprävention)
- Gesundheitsrelevanter Lebensstil und Gesundheitsverhalten
- Lebens- und Umweltbedingungen
- Soziodemografie und Sozialstatus
- Thematische Schwerpunkte sind weiterhin chronische Krankheiten, gesundheitliche Folgen des demografischen Wandels (steigender Anteil älterer Personen an der Bevölkerung) und psychische Gesundheit.

Stichprobenziehung, Studienpopulation

- Zielpopulation von DEGS1 waren die in Deutschland im Erhebungszeitraum lebenden und in den Einwohnermeldeämtern mit Hauptwohnsitz gemeldeten Erwachsenen im Alter von 18-79 Jahren. Um diese Grundgesamtheit zu repräsentieren, wurde eine Stichprobe gezogen.
- Die Studienpopulation besteht einerseits aus Personen (18-79 Jahre), die neu in die Untersuchung einbezogen wurden, zum anderen aus Personen (28-91 Jahre), die schon am BGS98 teilgenommen hatten und zu einer Teilnahme an DEGS1 bereit waren (Mischdesign). Die neuen Studienteilnehmer wurden über die Einwohnermeldeämter nach einem Zufallsverfahren ausgewählt und zur Untersuchung eingeladen.
- Die Untersuchungszentren befanden sich an 180 Studienorten im gesamten Bundesgebiet. Darunter waren alle 120 Studienorte des BGS98 sowie 60 weitere Studienorte, die zusätzlich ausgewählt wurden.
- Die Teilnehmer von DEGS1 bilden die Basis für zukünftige Nachfolgeerhebungen (Befragungen und ggf. Untersuchungen).

Stand: 22.05.2013

Untersuchungsteile und -ablauf	▪ Das Erhebungsprogramm beinhaltete eine schriftliche Befragung der Studienteilnehmer, körperliche Untersuchungen und Tests, ein ärztliches Interview, ein Arzneimittelinterview und Laboruntersuchungen von Blut- und Urinproben. ▪ In der medizinischen Untersuchung wurden u. a. Körpergewicht und -größe, Blutdruck und Puls, Schilddrüsengröße und Aspekte körperlicher Fitness erfasst. ▪ Durch Blut- und Urinanalysen sollen gesundheitliche Risiken festgestellt werden, die über eine Befragung und die körperliche Untersuchung nicht erkannt werden können, wie bspw. Nährstoffmangel, allergische Sensibilisierung oder Risiken für spätere Herz-Kreislauf-Erkrankungen. ▪ Die Studienteilnehmer wurden zum Ende ihres Untersuchungstermins in einem ärztlichen Abschlussgespräch über erste Befunde informiert und erhielten ca. sechs Wochen nach der Untersuchung einen schriftlichen Befund mit den gemessenen Laborparametern. Bei auffälligen Befunden wurde ihnen empfohlen, einen Arzt aufzusuchen. ▪ BGS98-Teilnehmer, die diesmal kein Untersuchungszentrum besuchen wollten oder konnten (z. B. weil sie inzwischen in einen anderen Ort verzogen waren), hatten die Möglichkeit, alternativ ein Befragungsprogramm (Gesundheitsfragebogen und telefonisches ärztliches Interview, keine Untersuchungen) zu absolvieren.
Modulstudie	▪ Von September 2009 bis März 2012 wurde die Kernstudie durch die „Zusatzuntersuchung psychische Gesundheit" ergänzt. In Befragungen wurden Daten zu beispielsweise Depressionen und den Umgang mit belastenden Lebensereignissen erhoben. Die Modulstudie wurde im Auftrag des RKI durch das Institut für Klinische Psychologie und Psychotherapie an der Technischen Universität Dresden durchgeführt.
Beteiligung	▪ Insgesamt nahmen 8.152 Frauen und Männer im Alter von 18-91 Jahren an DEGS1 teil. Darunter waren 7.238 Personen, die in den 180 Studienorten sowohl das Untersuchungs- als auch das Befragungsprogramm absolvierten, sowie weitere 914 Personen, die am Befragungsprogramm teilnahmen. ▪ Die Response betrug 62% bei den wiedereingeladenen und 42% bei den erstmals eingeladenen Personen.
Repräsentativität/ Gewichtung	▪ Die DEGS1-Daten ermöglichen für den Altersbereich von 18-79 Jahren (7.988 Personen, davon 7.116 in Untersuchungszentren) für Deutschland repräsentative Querschnittsanalysen und Trendaussagen im Vergleich mit dem BGS98. ▪ Die Querschnitts- und Trendanalysen werden mit einem Gewichtungsfaktor durchgeführt, der Abweichungen der Stichprobe von der Bevölkerungsstruktur (Stand 31.12.2010) hinsichtlich Alter, Geschlecht, Region und Staatsangehörigkeit sowie Gemeindetyp und Bildung korrigiert. Für den Untersuchungsteil wurde ein gesonderter Gewichtungsfaktor erstellt. Bei der Berechnung der Gewichtung für die ehemaligen Teilnehmenden des BGS98 wurde die Wiederteilnahmewahrscheinlichkeit berücksichtigt. Für die Durchführung von Trendanalysen werden die Daten des BGS98 auf den Bevölkerungsstand zum 31.12.2010 altersadjustiert.
Ergebnisse/ Publikationen	▪ Erste DEGS1-Ergebnisse wurden im Juni 2012 auf einem Symposium vorgestellt. ▪ Eine Informationsbroschüre für die Teilnehmerinnen und Teilnehmer der Studie mit Artikeln und Interviews zu ausgewählten Themen erschien Ende 2012. ▪ Auswertungsergebnisse zu einem breiten Themenspektrum wurden Ende Mai 2013 in einem Doppelheft des Bundesgesundheitsblattes veröffentlicht (Basispublikation). ▪ Die Daten werden ab 2014 in anonymisierter Form als „Public Use File" für die Gesundheitsforschung, Epidemiologie und Fachöffentlichkeit zur Verfügung gestellt.
Datenschutz und Ethik	▪ Die Teilnahme an der Untersuchung oder einzelnen Untersuchungsteilen war freiwillig. Die erhobenen Daten werden anonymisiert ausgewertet. ▪ Dem Bundes- sowie den Landesbeauftragten für den Datenschutz wurde das Studienkonzept vorgelegt. Die gegebenen Hinweise wurden berücksichtigt. ▪ Die zuständige Ethikkommission der Charité – Universitätsmedizin Berlin hat die Studie unter ethischen Gesichtspunkten geprüft und ihr zugestimmt.
Wissenschaftliche Begleitung	▪ Die Durchführung und Auswertung der gesamten Studie wird von der RKI-Kommission „Gesundheitsberichterstattung und Gesundheitsmonitoring" wissenschaftlich begleitet. ▪ Die Studie orientiert sich an den Leitlinien zur Sicherung „Guter Epidemiologischer Praxis" der Deutschen Gesellschaft für Epidemiologie (DGEpi), deren Einhaltung durch eine interne und externe Qualitätssicherung während der Studie kontrolliert wird.

Appendix X: Declaration of academic honesty

Hereby, I declare that I have composed the presented master thesis independently on my own and without any other resource than the ones indicated. All thoughts taken directly or indirectly from external sources are properly denoted as such.

Hamburg, 26th April 2021

____

Cedric, Noune Tankou