**Health Survey for England hypertension analysis and local prevalence models**

**Project for Public Health England**

**Miguel Pugliese, Roger Newson, Bowen Su, Michael Soljak**

**Department Primary Care & Public Health**

**School of Public Health**



**Contents**

[1 Background 4](#_Toc455842259)

[1.1 Epidemiology 5](#_Toc455842260)

[1.2 Risk factors 7](#_Toc455842261)

[1.2.1 Risk factor – Age 7](#_Toc455842262)

[1.2.2 Risk factor – Ethnic group 7](#_Toc455842263)

[1.2.3 Risk factor – Body mass index (BMI) 8](#_Toc455842264)

[1.2.4 Risk factor – Socioeconomic status 8](#_Toc455842265)

[1.2.5 Risk factor – Salt intake 8](#_Toc455842266)

[1.2.6 Risk factor – Reduced Potassium intake 8](#_Toc455842267)

[1.2.7 Risk factor – Physical activity 8](#_Toc455842268)

[1.2.8 Risk factor – Alcohol 9](#_Toc455842269)

[1.2.9 Other risk factors 9](#_Toc455842270)

[1.3 Treatment 11](#_Toc455842271)

[1.4 Previous hypertension prevalence modelling 11](#_Toc455842272)

[3 Methods 12](#_Toc455842273)

[3.1 Data source & sampling 12](#_Toc455842274)

[3.2 Outcome variables 12](#_Toc455842275)

[3.3 Risk factors 13](#_Toc455842276)

[3.4 Descriptive analyses 13](#_Toc455842277)

[3.5 Missing values 13](#_Toc455842278)

[3.6 Regression modelling 13](#_Toc455842279)

[3.7 Internal validation 14](#_Toc455842280)

[3.8 Local prevalence estimates: data sources 15](#_Toc455842281)

[3.9 Local prevalence estimates: Inverse probability weights method 15](#_Toc455842282)

[3.10 “External” validation of local estimates 16](#_Toc455842283)

[4 Results 17](#_Toc455842284)

[4.1 Population and baseline characteristics 17](#_Toc455842285)

[4.2 Statistical analyses 19](#_Toc455842286)

[4.2.1 Descriptive Statistics for the final population 19](#_Toc455842287)

[4.2.2 Missing values 22](#_Toc455842288)

[4.2.3 Univariable analyses 22](#_Toc455842289)

[4.2.4 Multivariable analyses 25](#_Toc455842290)

[4.3 Internal validation 28](#_Toc455842291)

[4.3.1 Tests for interactions and collinearity 28](#_Toc455842292)

[4.4 Discrimination 30](#_Toc455842293)

[4.4.1 ROC curves 30](#_Toc455842294)

[4.4.2 Predicted probabilities of having hypertension 31](#_Toc455842295)

[4.4.3 Sensitivity and specificity analysis 31](#_Toc455842296)

[4.5 Local estimates 32](#_Toc455842297)

[4.5.1 Internal validation of local estimates 34](#_Toc455842298)

[4.5.2 External validation of local estimates 34](#_Toc455842299)

[5 Discussion 40](#_Toc455842300)

[5.1 Summary 40](#_Toc455842301)

[5.2 Strengths and limitations 40](#_Toc455842302)

[5.2.1 Strengths 40](#_Toc455842303)

[5.2.2 Limitations 40](#_Toc455842304)

[6 References 41](#_Toc455842305)

[1. Annex 1: ROC curves 49](#_Toc455842306)

**HSfE Hypertension Analysis for PHE:**

**Technical Document**

# Background

Blood pressure is a normally distributed biological variable in which values have a continuous and graded relation with the risk of various cardiovascular diseases. Within this range of values, most national and international guidelines share a common threshold for hypertension, defined as a persistently raised blood pressure above an arbitrary cut-off point of ≥140 mm Hg for systolic blood pressure and ≥90 mm Hg for diastolic blood pressure.[1] Defined as primary or secondary hypertension, the former will be a combination of genetic (over 30%) and environmental factors whilst the latter will be symptoms caused by a clear cause, such as a iatrogenic trigger or renal or endocrine disorders, and with no family history. [1]

According to current the National Institute for Health and Care Excellence (NICE) guidelines, [2] hypertension is defined as one of the following;

* Stage 1 hypertension clinic blood pressure is 140/90 mmHg or higher and subsequent ambulatory blood pressure monitoring (ABPM) daytime average or home blood pressure monitoring (HBPM) average blood pressure is 135/85 mmHg or higher.
* Stage 2 hypertension clinic blood pressure is 160/100 mmHg or higher and subsequent ABPM daytime average or HBPM average blood pressure is 150/95 mmHg or higher.
* Severe hypertension clinic systolic blood pressure is 180 mmHg or higher or clinic diastolic blood pressure is 110 mmHg or higher.

In order to measure blood pressure monitoring to determine hypertension, the following procedure is recommended by NICE; [2]

1. When considering a diagnosis of hypertension, measure blood pressure in both arms.

* If the difference in readings between arms is more than 20 mmHg, repeat the measurements.
* If the difference in readings between arms remains more than 20 mmHg on the second measurement, measure subsequent blood pressures in the arm with the higher reading.

1. If blood pressure measured in the clinic is 140/90 mmHg or higher:

* Take a second measurement during the consultation.
* If the second measurement is substantially different from the first, take a third measurement.
* Record the lower of the last two measurements as the clinic blood pressure.

1. If the clinic blood pressure is 140/90 mmHg or higher, offer ambulatory blood pressure monitoring (ABPM) to confirm the diagnosis of hypertension.
2. If a person is unable to tolerate ABPM, home blood pressure monitoring (HBPM) is a suitable alternative to confirm the diagnosis of hypertension.

In spite of being both diagnosed as hypertension, high systolic and diastolic blood pressure are more strongly associated with different cardiovascular events; while the former is associated with angina, myocardial infarction and peripheral arterial disease, the latter is with abdominal aortic aneurism. [3 128] Moreover, data from different cohorts suggests that some events such as stroke were better predicted by episodic changes in systolic blood pressure compare to continued high mean blood pressure. [1]

It is also worth noting that the concept of “prehypertension” is gaining credence in the USA. The Seventh Report of the USJoint National Committee (JNC) on Prevention, Detection, Evaluation,and Treatment of High Blood Pressure provides a new classificationof blood pressure levels. [4] According to the new report,normal BP is defined as systolic BP (SBP) less than 120 mm Hgand diastolic BP (DBP) less than 80 mm Hg; a SBP of 120 to 139mm Hg or a DBP of 80 to 89 mm Hg is defined as prehypertension.The low-end threshold for prehypertension is lower than theprevious designation of high-normal BP (i.e. SBP/DBP: 130/85. However there islittle separation between the distributions of the risk factorsin people who over a specified period do or do not have a diseaseevent. With such closely overlapping distributions there areno cut-off levels that include most people who will have diseaseevents but few of those who will not have them, but rather there is a spectrum of risk as BP increases. [5] In addition, management of overall CVD risk rather than managing individual risk factors is now recommended. The definitions used by NICE and the HSfE have therefore been used for this model.

## Epidemiology

Hypertension is an important public health challenge in both economically developing, and developed, countries. An international review estimated that 26·4% of the adult population in 2000 had hypertension. [6] The estimated total number of adults with hypertension in 2000 was 972 million, of which 333 million are in in economically developed countries. The number of adults with hypertension in 2025 was predicted to increase by about 60% to a total of 1·56 billion. In 2010, high blood pressure was the biggest single contributor to the worldwide burden of disease. [7] However, some high-income countries, including Canada and England, have recently experienced big improvements in terms of awareness, treatment and control of hypertension. [8 ,9]

The effects of hypertension on the population are mainly driven by its contribution to cardiovascular diseases (CVD), including stroke and coronary heart disease (CHD). The relative risk for this two events has been observed to be similar in men and women. [10] However, hypertension is also a major risk factor for conditions such as arrhythmia, heart failure and renal disease, and there is increasing evidence of its central role as vascular risk factor for the development and progression of cognitive decline and dementia. [11] In addition, half of the disease burden attributed to high blood pressure is related to values below the commonly agreed 140/90 cut-off point. [12]Throughout middle and old age, usual blood pressure is strongly and directly related to vascular (and overall) mortality, without any evidence of a threshold down to at least 115/75 mm Hg.[13] Therefore, for most countries in the world, over 80% of all adults are at risk of CVD from their BP, with a large number of individuals and deaths concentrated in the 130/90 mm Hg range. This makes population-based approaches through diet and lifestyle, aimed at achieving a downward shift in the distribution of blood pressure in the whole population, an effective approach to reduce the burden of CVD. [14]

Numerous national and regional studies have been conducted worldwide to estimate the prevalence of hypertension: hypertension rates have varied from 3.4% in rural Indian men to as high as 72.5% in Polish women.[15 ,16] Within high income countries, average BP is higher in Europe than North America, even adjusting for treatment differences.[17] A meta-analysis of national studies showed a mean BP of 136/83 mm Hg in six European countries and 127/77 mm Hg in Canada and the United States among men and women combined who were 35 to 74 years of age. This difference already existed among younger persons (35-39 years) in whom treatment was uncommon (i.e. 124/78 mm Hg and 115/75 mm Hg, respectively), and the slope with age was steeper in the European countries. For all age groups, BP measurements were lowest in the United States and highest in Germany. The age and sex-adjusted prevalence of hypertension was 28% in the North American countries and 44% in the European countries at the 140/90 mm Hg threshold. Lower treatment thresholds and more intensive treatment contribute to better hypertension control in the United States compared with the western European countries (including the UK) studied. [18]

Table 1 Prevalence estimates of hypertension from literature

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Publication[[1]](#footnote-1) | Country | Type of study | Years covered | Total\* | Male\* | Female\* | Population age (years) |
| Jacob et al. | Germany | Cross-sectional, analysis of secondary GP data 2014 | 2014 | 66% | 65.7% | 66.1% | >65 |
| Moradi Lakeh et al.[19] | Saudi Arabia | Cross-sectional survey | 2013 | - | 4.36% | 2.46% | 15-24 |
| Wu et al. [20] | China | Cross-sectional, national survey | 2007-2011 | 24.3 [23.7-24.8] | 27.4% [26.6-28.2] | 21.6% [20.9-22.3] | ≥18 |
| Yoon et al.[21] | United States | Cross-sectional, national survey 2011-2014 | 2011-2014 | 29% | 30.0% | 28.1% | ≥18 |
| Joffres et al.[22] | Canada | Cross-sectional, national surveys | 2007-2009 | 19.5% | 19.7% | 19.3% | 20-80 |
| Joffres et al.[23] | England | Cross-sectional, national surveys (HSfE) | 2006 | 30.0% | 32.9% | 27.3% | 20-80 |

A recent study comparing national surveys from England, Canada and the United States found that, among these three countries, England had the highest hypertensions prevalence levels as well as the lowest levels of awareness, treatment and control.[23] The same way, a strong relation was found between these indicators and stroke and CHD. In Canada, 83% participants were aware of being hypertensive, 80% were treated and 66% treated and controlled. [23] Similarly, 54% of adults had controlled hypertension in the US in 2014 [21]. An exception in Western Europe might be France, where a study found that more than two-thirds of those with hypertension were aware of their diagnosis and 81% were treated with antihypertensive drugs. [24] In comparison, only 65% of English hypertensive patients were aware of their condition in 2006. From these, only 51% were controlled.

Within the United Kingdom, Macdonald and Morant[25] compared the prevalence of diagnosed and treated hypertension for the years 1998, 2003 and 2006 using the Health Survey for England and a general practice-based database; The Health Improvement Network (THIN). Hypertension was defined by either a recorded data analysis, an average of three successful blood pressure recordings over the 140/90 threshold or a record of antihypertensive drug prescribing. Overall, prevalence of hypertension was higher in the surveyed data compared to the THIN database during 1998, 37.3% to 25.3%, and 2003, 33% to 28%. According THIN data, treatment improved in patients from 45.2% in 1998 to 60.3%, however, it was not specified the percentage with controlled hypertension and low prevalence levels suggested a number unidentified cases. [25]

Table 2 Prevalence of hypertension by age and sex (shown as percentages) in England (HSfE and THIN), Macdonald and Morant[25]

|  | Female | | | | | | | | Male | | | | | | | | Both |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Reference | **16-24** | **25-34** | **35-44** | **45-54** | **55-64** | **65-74** | **≥75** | **All ages** | **16-24** | **25-34** | **35-44** | **45-54** | **55-64** | **65-74** | **≥75** | **All ages** | **All ages** |
| (1998 -HSfE) | 4.3 | 6.6 | 13.1 | 30.9 | 49.3 | 72.2 | 77.8 | 33.5 | 19.0 | 19.3 | 26.0 | 41.3 | 60.1 | 70.2 | 73.6 | 41.6 | 37.3 |
| (1998 -THIN) | 2.6 | 6.3 | 12.6 | 29.1 | 48.0 | 58.8 | 59.9 | 27.8 | 1.6 | 4.6 | 11.0 | 24.6 | 41.1 | 54.6 | 57.2 | 22.6 | 25.3 |
| (2003 -HSfE) | 1.4 | 5.3 | 10.2 | 22.8 | 43.2 | 63.5 | 75.0 | 31.0 | 6.7 | 11.6 | 17.8 | 35.4 | 47.7 | 62.5 | 68.1 | 35.2 | 32.9 |
| (2003 -THIN) | 3.3 | 7.5 | 14.3 | 31.1 | 50.5 | 62.9 | 63.7 | 30.1 | 1.7 | 5.2 | 12.0 | 26.6 | 44.6 | 60.7 | 62.8 | 25.3 | 27.8 |
| (2006 -THIN) | 2.7 | 7.1 | 13.7 | 28.0 | 48.1 | 60.7 | 62.2 | 28.9 | 1.4 | 4.6 | 10.9 | 24.7 | 43.4 | 59.8 | 61.7 | 24.7 | 26.9 |

## Risk factors

For this project, we conducted a non-systematic literature search on the different risk factors affecting hypertension according to the most recent literature available (Table 3). We have summarized the evidence from different sources in this section, by risk factor, as well as reflected the debate present in a few of them (Table 4).

Table 3: Hypertension risk factor list

|  |  |
| --- | --- |
| Risk factor | References |
| Age | [26-28] |
| Ethnic group | [26 ,29-33] |
| Obesity/BMI | [26 ,34] |
| Socioeconomic status | [28 ,34-42] |
| Salt intake | [43-45] |
| Reduced potassium intake | [14 ,46-48] |
| Physical activity | [49-51] |
| Alcohol intake | [52-55] |
| Air pollution | [56] |
| Depression | [57-59] |

### Risk factor – Age

Along with age, there is a steady increase in the prevalence of hypertension, after controlling for other variables such as Body mass index, socioeconomic status, marital status and physical activity. [26-28] During middle and old age, usual blood pressure is more strongly related to vascular mortality, without any evidence of a threshold to at least 115/75mm Hg. [60] Similarly, extensive data from the UK suggests that age might affect differentially the effect of systolic and diastolic blood pressure on different cardiovascular events. [3] In addition, there is strong evidence from diverse populations showing that blood pressure during childhood is associated with BP in later life, and early intervention is important. [28 ,34 ,61 ,62]

### Risk factor – Ethnic group

Differences in the probability of suffering hypertension between ethnic groups have been reported by studies performed in the UK and abroad, with individuals of African descent having higher odds than those of European descent. [26 ,29] However, differences in blood pressure measurement techniques, as well as differences in sex, age and BMI where acting as confounders. [29] Although some studies also point to greater chances of hypertension among individuals of South Asian descent, [26 ,31] evidence regarding this group is mixed, [30 ,33 ,63] being one of the possible reasons the aggregation of groups (Indian, Bangladeshi and Pakistani) in which blood pressure tends to vary in comparison to white individuals. [30] A recent meta-analysis grouping a total of 23 European studies—18 from the UK—concluded that men and women from Sub Saharan descent presented higher values of systolic and diastolic blood pressure than those from European descent. On the other hand, men and women of South Asian descent presented lower SBP and similar DBP values than their European counterparts. [32]

### Risk factor – Body mass index (BMI)

Several studies have observed a correlation between greater odds of suffering from hypertension and a BMI>25. [26 ,34] According to a cohort study conducted by Forman et al, BMI was the most powerful predictor of hypertension in women aged 27 to 44.[64] A linear association between BMI and blood pressure has been suggested [65] and it is been suggested that it would explain part of the differences observed in other variables, such as socioeconomic status. [35] In their meta-analysis, Arabshahi et al. suggest that the relation between BMI and hypertension might be mediated by other factors such as maternal malnutrition during pregnancy, varying the degree to which different individuals are exposed to suffer from increased blood pressure.[66]

### Risk factor – Socioeconomic status

In high income settings, socioeconomic status has been associated with the greater risk of developing hypertension since long. [28 ,34-36] This association has been observed in terms of education. [28 ,37] socioeconomic quintile [38] and economic trajectory. [39] Moreover, it is believed to explain part of the differences in hypertension observed between ethnic groups [40] and has been observed to modify relation between age and blood pressure, being the effect of age more acute in lower socioeconomic groups. [41] Similarly, certain job strains have been observed to be more associated with higher levels of hypertension, being higher job decision latitudes correlated in general with lower prevalence of hypertension with some exceptions such as healthcare support occupations.[42]

### Risk factor – Salt intake

The relation between salt and blood pressure has been known for long time; the INTERSALT study, a worldwide epidemiologic study including over 10,000 individuals already found that higher sodium intakes resulted in higher levels of systolic and diastolic blood pressure regardless of sex, age group and hypertension status. These findings were in line with epidemiological, clinical and animal experimental evidence showing a direct relationship between dietary electrolyte consumption and blood pressure. [43] More recently, systematic reviews and meta-analysis concluded that reducing sodium intake contributed to lower blood pressure among adult and children, with even small reductions in consumption over a month-long period producing significant reductions in both normotensive and hypertensive individuals. [44 ,67] Within England, an analysis of the Health Survey for England 2003-2011 data found evidence of a reduction in blood pressure in the non-hypertensive English population after adjusting for other variables, suggesting this is likely to be related with a reduction in salt consumption. [45]

Some large observational studies supporting a heterogeneous association of sodium/potassium consumption with blood pressure, being more pronounced in older individuals and those with hypertension and high levels of sodium consumption. [68 ,69] Phenotypic traits associated with an increased sensitivity to blood pressure are also likely to include individuals of African origin, suffering from obesity and with metabolic syndromes. [68]

### Risk factor – Reduced Potassium intake

A meta-analysis of 22 randomized controlled trials concluded in 2013 that increased potassium intake contributed to the reduction of blood pressure in people suffering from hypertension. [46] This reduction is additive to the effects produced by lower sodium-consumption, [47] is appreciated through an increased consumption of fruits and vegetables [48] and may, in certain circumstances, palliate the effects of high sodium intake. [14]

### Risk factor – Physical activity

The effects of physical activity on blood pressure are well documented through different studies. In their meta-analysis of 93 randomized controlled trials, Cornelissen and Smart found an association between different types of training and a reduction in systolic and diastolic pressure, with pre and hypertensive individuals experiencing bigger reductions than those participants with normal blood pressure. [49] In the Australian Longitudinal Study on Women, authors observed that the risk for hypertension decreased with increases in the total number of hours of metabolic time spent doing exercise, with little difference between moderate and vigorous activity,[51] similarly, a dose/response association was found among usual cyclists in the UK, showing greater reductions beyond the physical activity’ recommendations. [50]

### Risk factor – Alcohol

The relation between alcohol consumption and higher blood pressure is well documented in high income countries, where different studies have estimated that alcohol accounts for between 5 and 30% of the total prevalence of hypertension.[70] [52] The type and intensity of this relation has been observed to be dose-dependent, and vary between genders [53 ,54] and individuals of ethnic minority background. [53 ,71] However, some studies have observed large reductions in blood pressure shortly after withdrawal; according to their authors, this fast and progressive reduction challenge the assumption that alcohol produces a sustained increases of blood pressure in the long-term, being observed differences a result of recent consumption. [55] [70]

### Other risk factors

Other risk factors include air pollution, having been observed that particulate matters (diameter <2.5 µm) are associated with higher blood pressure, especially with long-term exposure.[56] Similarly, some studies suggest that depression is a risk factor for increased blood pressure, being this association time-dependent [57] and of varying intensity depending on gender [58 ,59] and age group. [59]

Although there is some debate with regards to the relation between smoking status and hypertension, with some certain observing current smokers to be at risk of higher systolic blood pressure, [72 ,73] there is strong evidence, including a Mendelian randomization meta-analysis with over 140,000 participants from 23 different studies, showing that smoking might increase resting heart rate, but not blood pressure. [74-76]

Table 4: Risk factors and their ORs from various studies

| Risk factor | Type of Odds Ratio | Odds Ratio | 95% CI | Effect on Outcome |
| --- | --- | --- | --- | --- |
| Age (years) |  |  |  |  |
| 20-39 | Adjusted | 1.0 |  | Reference |
| 40-59 | Adjusted | 8.3 | [3.4-20.3] | Risk factor |
| 60-79 | Adjusted [26] | 31.2 | [12.3-79.5] | Risk factor |
| Ethnic group (for males) |  |  |  |  |
| White | Adjusted | 1.00 |  | Reference |
| Afro-Caribbean | Adjusted | 1.56 | [1.14-2.13] | Risk factor |
| South Asian | Adjusted | 1.31 | [0.88-1.97] | Non-significant |
| Ethnic group (for females) | Adjusted |  |  |  |
| White | Adjusted | 1.0 |  | Reference |
| Afro-Caribbean | Adjusted | 2.4 | [1.51-3.81] | Risk factor |
| Body Mass index (BMI) |  |  |  |  |
| <25 | Adjusted | 1.0 |  | Reference |
| 25-30 | Adjusted | 2.0 | [1.2-3.3] | Risk factor |
| >30 | Adjusted | 3.5 | [2.2-5.6] | Risk factor |
| Education |  |  |  |  |
| Less than High School | Adjusted | 1.00 |  | Reference |
| High School graduate | Adjusted | 1.30 | [1.2-3.3] | Risk factor |
| More than High School | Adjusted [28] | 1.25 | [1.2-3.3] | Risk factor |
| Annual household income ($) |  |  |  |  |
| ≥ 50,000 | Adjusted | 1.00 |  | Reference |
| <15,000 | Adjusted [77] | 1.17 | [1.13–1.21] | Risk factor |
| 15,000-24,999 | Adjusted | 1.12 | [1.09–1.15] | Risk factor |
| 25,000-34,999 | Adjusted | 1.10 | [1.07–1.13] | Risk factor |
| 35,000-49,999 | Adjusted | 1.07 | [1.04–1.09] | Risk factor |
| Moderate physical activity (metabolic equivalent minutes) MET·min·wk-1 | [51] |  |  |  |
| 0 | Adjusted | 1.00 |  | Reference |
| 0 to 250 | Adjusted | 0.92 | [0.83-1.02] | Non-significant |
| 250 to <500 | Adjusted | 0.90 | [0.81-1.00] | Non-significant |
| 500 to <1000 | Adjusted | 0.82 | [0.75-0.91] | Protective factor |
| 1000 to <1500 | Adjusted | 0.74 | [0.66-0.82] | Protective factor |
| 1500 to <2000 | Adjusted | 0.78 | [0.68-0.90] | Protective factor |
| ≥ 2000 | Adjusted | 0.80 | [0.70-0.93] | Protective factor |
| Alcohol (grams per day) g/d |  |  |  |  |
| Men |  |  |  |  |
| Non-drinker | Adjusted | 1.00 |  | Reference |
| <10 | Adjusted | 1.03 | [0.94-1.13] | Non-significant |
| 11 to 20 | Adjusted | 1.15 | [0.99-1.33] | Non-significant |
| 21 to 30 | Adjusted [54] | 1.07 | [0.86-1.34] | Non-significant |
| 31 to 40 | Adjusted [54] | 1.77 | [1.39-2.26] | Risk factor |
| 41 to 50 | Adjusted [54] | 1.17 | [0.84-1.65] | Risk factor |
| >50 | Adjusted [54] | 1.61 | [1.38-1.87] | Risk factor |
| Women |  |  |  |  |
| Non-drinker | Adjusted [54] | 1.00 |  | Reference |
| <10 | Adjusted [54] | 0.87 | [0.82-0.92] | Protective factor |
| 11 to 20 | Adjusted [54] | 0.90 | [0.87-1.04] | Non-significant |
| 21 to 30 | Adjusted [54] | 1.16 | [0.91-1.46] | Non-significant |
| 31 to 40 | Adjusted [54] | 1.19 | [1.07-1.32] | Risk factor |
| Depression (increase per 5 years) |  |  |  |  |
| Low /transient depression | Adjusted [58] | 1.00 |  | Reference |
| Increasing depression | Adjusted [58] | 1.07 | [1.03-1.12] | Risk factor |

## Treatment

Drug treatment is recommended in all people with sustained levels of blood pressure ≥160/100, and at lower levels in those with either additional risk factors or target organ damage. The threshold for offering drug treatment is defined by NICE [2] as:

* blood pressure of more than 160/100 mmHg (hypertension stage 2) or
* blood pressure of more than 140/90 mmHg and:
  + target organ damage
  + established cardiovascular disease
  + renal disease diabetes
  + a 10-year cardiovascular risk equivalent to 20% or greater. If only this factor is present and the patient is aged under 40, specialist evaluation of secondary hypertension causes and organ damage assessment are recommended.

## Previous hypertension prevalence modelling

In the US, self-reported data on hypertension diagnosis from the Behavioral Risk Factor Surveillance System was used to obtain unbiased state-level estimates of blood pressure and uncontrolled hypertension as benchmarks for priority setting and for designing and evaluating intervention programmes. [78] A crude epidemiologic hypertension model based on age and sex-specific prevalence was first developed in the UK by the Faculty of Public Health.[79] Subsequently, English primary care trusts (PCTs) were required to set targets for hypertension case-finding in their 2007-8 Local Delivery Plans negotiated with strategic health authorities. To assist them a comparatively simple PCT-based prevalence model was developed rapidly by the Association of Public Health Observatories (APHO). [80]

In the 2007 APHO model numbers of persons predicted to be hypertensive were derived by multiplying April 2006 PCT registered populations by hypertension prevalence rates identified in the 2003 and 2004 HSfE, modified by ethnic-group age-standardised risk ratios from the 2004 HSfE. Calculations were stratified to reflect variations in these factors at PCT level. The PCT-registered populations were derived by aggregating April 2006 GP practice populations from the Exeter System using a practice to a new PCT lookup table. In the absence of age by sex by ethnic-group PCT populations, the age by sex registered populations of the current PCTs were attributed the ethnic-group distributions of their constituent former PCTs resident populations from the 2001 Census.

APHO accepted that the 2007 model was rather outdated, and that a new prevalence model based on a comprehensive regression model using recent HSfE data would be more robust. This model was commissioned from the Department of Primary Care & Social Medicine at Imperial College London, and developed during 2008/09, using information from the Health Surveys for England 2003 and 2004.

# 

# Methods

## Data source & sampling

The hypertension model described here uses data from the 2013 and 2014 Health Survey for England (HSfE). The 2013 and 2014 data consists in two individual level files: hse2013ai (10,980 records) and hse2014ai (10,080 records) containing data for all individuals in the General Population Sample in co-operating households who gave a full interview. These two data files were merged for the purpose of this project.

For both surveys, multi-stage stratified probability sampling design was used, with a two stage sampling. At the first stage, a random sample of primary sampling units (PSUs) based on postcode sectors was selected. Within each, a random sample of 16 postal addresses of private households was selected. All adults aged 16 years and over at each household were selected for the interview (up to a maximum of ten adults per household). Nurse visits were offered to all participants who were interviewed. Nurses took measured the blood pressure of those aged 5 and over [81 ,82].

## Outcome variables

To perform the blood pressure measurements, a nurse visit to the respondent household was arranged. Using Omron HEM 907 blood pressure monitors, nurses performed three blood pressure measurements at with one minute-pause intervals. [81 ,82] Respondents who ate, smoked, drank alcohol or participated in vigorous exercise up to 30 minutes before the nurse visit were indicated as such and discarded from further analysis. Other events limiting the measurement of blood pressure during the visit, such as being pregnant or not obtaining three valid measurements, were also noted and included in this variable. From these measurements, the mean of the second and third blood pressure measurements for systolic and diastolic blood pressure was used to determine the hypertensive status of the person. A final population of individuals aged 16 or more and with three valid blood pressure measurements (n=9,883) was selected for the study.

Prevalence of hypertension was defined as systolic blood pressure ≥140 or diastolic blood pressure ≥90, according to commonly used standards. The specific definitions of the four levels used in the HSfE 2013 report are:

* **Normotensive** SBP <140mmHg and DBP <90mmHg and not taking medicine prescribed for high blood pressure
* **Hypertensive controlled** SBP <140mmHg and DBP <90mmHg and taking medicine prescribed for high blood pressure (in 2003 this was called “normotensive-treated”)
* **Hypertensive uncontrolled** SBP ≥140mmHg and/or DBP ≥90mmHg and taking medicine prescribed for high blood pressure (in 2003 this was called “hypertensive-treated”)
* **Hypertensive untreated** SBP ≥140mmHg and/or DBP ≥90mmHg and not taking medicine prescribed for high blood pressure (in 2003 this was called “hypertensive-untreated”)

In this document, two different for models for hypertension, diagnosed and undiagnosed, are presented. Therefore two different outcome measures were obtained. The first outcome, diagnosed hypertension, was derived from the number of patients that reported being told by a nurse or doctor that they had high BP. Respondents who also reported being pregnant when they were informed were excluded unless they were also told in a different occasion. Those answering who didn’t know (21 respondents) or refused answering (1 respondent) whether they ever had been told high blood pressure, were considered as never being diagnosed for high blood pressure.

The second outcome, undiagnosed hypertension, was derived from those patients that, first, were considered uncontrolled or untreated hypertensive and second, they did not report having been diagnosed high blood pressure.

## Risk factors

We used the literature review described in the Background to extract HSfE data on risk factors. There were two main reasons why some risk factors from the literature were not used in the final model. Firstly, the data was not available in HSfE (eg air pollution and salt intake) or some form of data was available, but did not fit adequately the needs of the model. Data on depression would be an example of the latter, as only self-reported data on whether the respondent suffer from depression or anxiety (same variable) in the last twelve months was available. Given some contradictory evidence in the literature, cigarette smoking was also tested as a possible variable in order to check any possible effects in the model. We performed a number of descriptive analyses on the informant-level dataset including demographics, risk factor breakdowns and categories, and individuals suffering from diagnosed and undiagnosed hypertension.

## Missing values

There were a number of missing values in the survey, identified as respondents who either did not know or refuse to provide an answer to a specific question. IN the HSfE, all these values are coded with a different negative number depending on the reason why the answer was not obtained. In order to deal with missing values, the multiple imputation approach was used. Given the evident problems that complete-case analysis brings due to the deletion of incomplete observations, multiple imputation is a widely accepted technique to estimate those values missing.[83-85] Under the assumption that missing values on the dataset are missing at random (MAR), regression analysis is conducted on variables registered as imputed (i.e. with missing values) using other complete variables. Through this analysis, missing values will be inferred from the relation of the imputed variable with the other complete (regular) variables. This process will be conducted a minimum of five times, a mean of these n=5 analyses will be the imputed value substituting the missing information on each observation.

There are two exceptions where multiple imputation was not performed: respondents reporting not knowing whether they ever had high blood pressure were considered to never have been diagnosed by a doctor or nurse either. Secondly, individuals “not knowing” their ethnic origin were considered to be white.

## Regression modelling

The choice of variables for original inclusion in the merged dataset included all those known to be high blood pressure risk factors. The variable names and labels are shown in Table 3. The HSfE dataset has a nested or hierarchical structure so three variables related to the sampling strata were included: area (sample point), cluster (stratification level), and wt\_nurse (weights accounting for non-respondents and survey structure). These were used in the model to adjust for clustering of respondents.

We fitted univariate then multivariate logistic regression models for diagnosed and undiagnosed hypertension to produce odds ratios (ORs) and regression coefficients. Model specification was conducted using forward stepwise selection and each newly introduced variable was tested using Wald tests. A range of multivariate regression models were fitted in order to obtain the best performing.

The modelling and estimation of the effects of interest was carried out using the logit command. The initial output consisted of two tables: one with the estimated regression coefficients, corresponding p-values and 95% confidence intervals, and another with the estimated odds ratios (exp(b)), which in the table appear as relative risk ratios (RRRs) and 95% confidence intervals. A positive sign of the estimated coefficient is associated with an increase in the odds of the outcome had diagnosed/undiagnosed hypertension, and a negative sign is associated with a decrease in the odds. Since Prob (A) = Odds (A) / 1+ Odds (A), for uncommon outcomes such as high blood pressure, RRR can be assumed to be the same as the odds ratio (OR).

For categorical variables the effects are estimated relative to the reference category. Stata uses the first category as reference (baseline OR). Separate baseline odds were estimated for each gender, and also according to ethnicity, age band, area-based deprivation score etc. The model can be used to derive the prevalence ratios for high blood pressure for subjects with various combinations of risk factors in relation to baseline. The prevalence in each age group, gender, ethnic group, area of residence and level of deprivation category were derived from the odds, using the formula: prevalence = odds/(1 + odds).

There is an interaction between the effects of two exposures if the effect of one exposure varies according to the level of the other exposure.[86] For example, there might be an interaction between the hypertension risk factors of education level and social class. An alternative term for interaction its effect modification. In this example, we can think of this as educational level modifying the effect of social class. For this model, we tested for interactions between HSfE predictor variables for risk factors. Wald tests were used to compare the possible interactions and decide whether they should be included.

## Internal validation

Ideally the best prediction should result from utilising the most information in the regression model. However only a limited range of HSfE variable data is either available or can be estimated at the clinical Commissioning Group (CCG) and general practice levels. Therefore, we decided to validate a local model (that only used locally available data) by comparing it, in terms of prediction, to a complete model including all available and significant HSfE variables. This was done separately for diagnosed and undiagnosed hypertension.

In order to use local Census and other data as sources for our estimates, changes were made in our local model in order to fit the data. These included dropping the alcohol consumption variable within the undiagnosed hypertension model and readjusting categories in both of them. A description of the changes is included in the “local model” section in the results.

Finally, we internally validated the models by generating receiver operating characteristic (ROC) curves, by using the ***predict*** regression post-estimation command to generate for each HSfE informant the probability of having hypertension using the derived odds ratios (ORs), and by using these probabilities to examine sensitivity and specificity.

All statistical analysis was carried out in StataSE14.

## Local prevalence estimates

### Data sources

Various sources were used for local risk factor data as follows:

* practice age/sex breakdowns were obtained from the Health & Social Care Information website: <http://www.hscic.gov.uk/searchcatalogue?topics=1%2fPrimary+care+services%2fGeneral+practice&sort=Most+recent&size=10&page=2#top>
* Top qualification from Census 2011 data available on the Neighbourhood Statistics website, mapped to practices: <http://www.neighbourhood.statistics.gov.uk/dissemination/Download1.do?&nsjs=true&nsck=false&nssvg=false&nswid=1920>
* Ethnicity: Office for National Statistics
* Employment category from Census 2011 data available on the Neighbourhood Statistics website, mapped to practices: <http://www.neighbourhood.statistics.gov.uk/dissemination/Download1.do?&nsjs=true&nsck=false&nssvg=false&nswid=1920>
* Index of Multiple Deprivation (IMD) mapped to practices: Department of Communities & Local Government: <https://www.gov.uk/government/collections/english-indices-of-deprivation>. IMD quintiles are not available nationally for any level, so we had to calculate the percentage of people in each quintile:
  + we divided Lower Super Output Areas (LSOAs) into quintiles according to their IMD score (using 2013-2014 criteria)
  + we converted them into populations, using the number of people contributed by each LSOA to each practice
  + we summed up for each practice IMD scores all the people in different LSOAs for each quintile
  + we turned these into percentage of the practice population in each IMDS quintiles.
* Body Mass Index:
* Limiting long-lasting illness from Census 2011 data available on the Neighbourhood Statistics website, mapped to practices: <http://www.neighbourhood.statistics.gov.uk/dissemination/Download1.do?&nsjs=true&nsck=false&nssvg=false&nswid=1920>

### Small population estimation methods

Derived ORs (or rather, regression coefficients) are used to estimate prevalence in small population subgroups. Local population breakdowns for each risk factor are used, where these are available. ICL has a wide range of small population risk factor prevalence breakdowns, including age, sex, deprivation, alcohol consumption, ethnicity, long-term illness, anxiety, BMI and marital status. The local model uses locally available data and so includes only those variables that are available at local population level i.e. age, sex, socioeconomic status, BMI, and other disease conditions. The steps in applying the prevalence estimates are as follows and in the equations below:

* Use the regression coefficients to generate log odds (since they are from a logistic regression model) for each risk factor subcategory
* Generate a similar table of odds by exponentiation
* Generate a similar table of prevalence in each risk factor subcategory using the epidemiologic formula
* Produce a matching table of small population subcategories. If there are no corresponding local data with a sufficiently granular breakdown e.g. ethnicity by age by sex, this requires deciding how each risk factor should be attributed across other risk factor categories, with evenly as the default. For example, we used the national age/sex/ethnicity breakdown from the Census and age breakdowns from the HSfE to attribute this data at small population levels. The actual breakdown will be somewhat different and needs to be borne in mind as another source of potential error.
* Multiply the population cells by the corresponding prevalence to estimate the number of people in each cell with the disease

In mathematical notation:

Predicted log odds of prevalence = *b0* + *b1x1i* +  *b2x2 i* + *b3x3 i* +  *b4x4 I*

where *b0* = regression constant, *b1, b2,  b3, b4*= other regression coefficients

*x 1 i, x2 i, x3 i, x4 i* = value of risk factors for individual ***i***

(NB since all the variables are binary variables, *x* =1 if specified risk factor is present, *x*=0 if it is absent). Predicted log odds of prevalence for a community of *n* individuals is derived by averaging over the values for all individuals included in the community:

Predicted log odds of prevalence in community of *n* individuals:

= 1/*n* ∑i=1n (*b0* + *b1x1i* +  *b2x2 i* + *b3x3 i* +  *b4x4 i)*

= *b0* + *b1p1* +  *b2p2* + *b3p3* +  *b4pp4*

where *p1 , p2, p3, p4*=proportion of individuals in the community with characteristic *x1 , x2 , x3 , x4 .* (i.e. proportion with *x.*=1 rather than *x.*=0 as in the remainder).

The predicted prevalence for an individual is derived from their predictive log odds using:

prevalence = exp(log odds)/[1+exp(log odds)]

=*exp*(*b0* + *b1x1i* +  *b2x2 i* + *b3x3 i* +  *b4x4 i) /[1+* *exp*(*b0* + *b1x1i* +  *b2x2 i* + *b3x3 i* +  *b4x4 i)]*

Predicted prevalence in community of *n* individuals:

= 1/*n* ∑i=1n{*exp*(*b0* +*b1x1i* +*b2x2 i* +*b3x3 i* +*b4x4 i)/[1+* *exp*(*b0* +*b1x1i* +*b2x2 i* +*b3x3 i* +*b4x4 i)]*

Unfortunately, the equation above does not simplify to a linear combination of the predictor variables (in the way the mean log odds does). The average/overall prevalence is not the same as the prevalence for a person with “average” risk factors. So, for instance, it cannot be found by taking exp(log odds)/[1+ exp(log odds)] of the average log odds. There is no linear relationship with the regression coefficients, and with proportions of population with specified risk factors.

In order to find a synthetic estimate of prevalence, ideally we need to know the joint distributions of the included risk factors in the relevant population (the population on which are synthetic estimates are required). Ideally, we would know how many people in the population have each specific combination of risk factors. In practice, it might be good enough to know the distribution of some risk factors individually, rather than in combination. For instance, we might know what proportion of the population are smokers, and what proportion are ex-smokers, but not how many smokers we have by age and sex. In this situation, we have assumed that the same proportion of all ages and both genders are smokers and ex-smokers. Even if this is not exactly correct, then the synthetic estimate of prevalence may still be a reasonably accurate estimate (assuming that the smoking distribution does not vary too much by age, sex and other included risk factors). This is considered a good enough approach, and the best possible based on the information currently available in many cases.

In practice, we know the population distributions by age and sex, therefore we do not need to make the assumption that the proportion of males is the same for each age group. We use the more precise method of using the actual proportions of males in each age group. From the ELSA longitudinal survey we also know that older people/ older females in particular are generally less educated (on the basis of qualifications held). Therefore we apply the proportions with any educational qualifications according to age and sex group.

For other risk factors, we do not know whether these risk factors are more or less common in males than in females, nor according to age group, nor educational status i.e. we do not know their distributions in combination with any of the other risk factors included in the model. Therefore we make the assumption that the distribution of all other risk factors (apart from afore-mentioned age, sex and educational status), is equal across all other risk factors. This makes the calculations somewhat easier, even though this assumption might make for slightly less accurate estimates, the loss of accuracy is not thought to be great.

In order to find the estimated prevalence for each population, it is necessary to calculate the synthetic prevalence of risk factors for each possible combination of risk factor (as included in the chosen disease-specific logistic regression model). The estimated prevalence for a population is then the weighted average of the prevalence estimates for each combination of risk factors, according to the estimated number of people with each risk factor combination in the population (the population on which synthetic estimates are sought). These calculations can be carried out in Excel (using VBA code to link prevalence and risk factor spreadsheets with formulae in a workbook) or in Stata software to produce confidence intervals as well as the estimates. We have developed two methods of producing these estimates, with CIs, in Stata. The first uses inverse probability weighting, the second using a “bootstrap” procedure. Ideally we wish to use both methods as an additional internal validation. However the short timeframe did not permit this so the estimates have been produced using inverse probability weighting only. More information about the bootstrap method is included in Annex 1.

### Bootstrap method for local prevalence estimates

In summary, within Stata, a new set of variables is created, one for each combination of these risk factors pertinent to the logistic regression model for the chosen disease. With our dataset set up in this way, we can now use Stata’s “*predict*” command to give us the predicted log odds. Then we find the weighted average of these, averaged across all possible combinations of risk factors, using the weights calculated as above (stored in variable named xyz). The weighted average can be found using the “collapse” command, which results in one line of data per practice or MLSOA (using the population identifier as the *by* variable) in Stata.

We have also calculated in Stata CIs for prevalence estimates using a “bootstrap” procedure. There is uncertainty in these synthetic estimates of prevalence based on the imprecision not in the more usual sample of people from the population (since the estimates are not a sample but are externally applied), but in the estimated coefficients from the logistic regression equations. A bootstrap procedure can be used to construct confidence intervals on these synthetic estimates of prevalence, based on the imprecision in these logistic regression coefficients.

The philosophy underlying the bootstrap procedure is to consider that the people included in the data set used to derive the logistic regression equation represent the whole population of possible people. However, the whole population is effectively considered to contain thousands of copies of each of these people. Bootstrap samples are taken randomly from our initial populations (the subsets of the HSfE population that has complete data on appropriate risk factors). Logistic regression of the same risk factors can then be applied to this boot strap sample, i.e. we rerun the logistic regression that gave us our chosen predictive model. However, we get slightly different regression coefficients, because of the modified sample. Prevalence estimates are then derived for each combination of risk factors, based on these new regression equations.

This process is repeated 1,000 times, to find 1,000 different boot strap samples, by random sampling processes, and to then fit logistic regression equations on each. The prevalence estimates are calculated for each combination of risk factors, for each of these 1,000 boot strap samples. For each small population, a synthetic estimate is calculated for each boot strap sample, by appropriately weighting the prevalence estimates on each combination of risk factors (with the same weights as described above which reflect the anticipated prevalence of each combination of risk factors in the population). From these 1,000 synthetic estimates of prevalence of each population, a 95% confidence interval is calculated as the 2.5th to 97.5th centiles. Given that the estimates are distributed normally, these are taken to be mean +/- 1.96 SD (taking mean and SD of the 1,000 boot strap synthetic prevalence estimates for each specified region).

### Local prevalence estimates: inverse probability weights method

Inverse probability weighting methods are used to standardize from a sampled population to a target population. They are usually defined as a function of a panel of one or more sampling-probability predictor variables. For each combination of the predictor variables, the sampling probability weight is the ratio of the frequency of that combination in the target population to the frequency of that combination in the sampled population. Inverse probability weighting is therefore a generalization of direct standardization. In Stata, it is implemented by using a *pweight* qualifier on an estimation command. This normally implies the use of a Huber variance formula to generate the confidence limits.

The “local” model includes only those variables that are available at local population level. Local population breakdowns for each risk factor are used, where these are available. ICL has a wide range of small population risk factor prevalence breakdowns, including age, sex, deprivation, education levels, employment category, BMI, long-lasting illness, ethnicity. The local model uses locally available data.

In Stata, we may apply inverse-probability weights to a regression model, using the *pweight* qualifier to standardize the regression results from a sampled population to a target population. Alternatively, once we have regression results from one set of data, we may use a *pweight* qualifier with a post-estimation command (such as margins or the add-on packages *margprev* or *marglmean* [87 ,88] to compute a predicted mean or prevalence, using out-of-sample prediction. As an example, we may compute predicted prevalence for practices, using *margprev* in a scenario dataset for each practice.

To do this, we save the estimation results for the regression model in a .ster file (containing Stata estimation results). For each practice, we create a scenario dataset, with 1 observation for each possible combination of predictor values in that dataset. The sampling probability weights are the presumed probabilities of each combination of predictor values in that practice. In practice, we do not know the combination probabilities for each practice, as we only have estimates of marginal probabilities. (Such as a distribution of gender/age, a distribution of smoking status, and a distribution of ethnic origin, without any combination or joint distribution.) So, we assume (for want of better knowledge) that predictors are statistically independent, and estimate combination probabilities from marginal probabilities (using the reshape long command in Stata). We then input the .ster file and the practice scenario datasets, and used the *margprev* add-on package, and the *parmby* module of the *parmest* add-on package, to estimate marginal prevalence of disease, using the model to make out-of-sample predictions for each practice to produce the synthetic estimates for that package. An example of this can be found in Newson (2014). [88] Confidence intervals (CIs) are also calculated for each local estimate by the packages.

## Validation of local estimates

### Internal validation of local estimates

The local estimates can also be validated by aggregating them to the lowest geography available in the raw HSfE data and comparing them. We also present these results.

When estimating the total effects of individual distal factors on disease, both mediated and direct effects should be considered, because in the presence of mediated effects, controlling for the intermediate factor would attenuate the effects of the more distal one. [89] When estimating the joint effects of the more distal factor and the intermediate one, the mediated and direct effects should be separated, especially if the intermediate factor is affected by other distal factors.

Finally, there can be collinearity between exposures to various risk factors, meaning that one can be linearly predicted from the others with a substantial degree of accuracy. In this situation the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data. Collinearity does not reduce the predictive power or reliability of the model as a whole, at least within the sample data set; it only affects calculations regarding individual predictors. That is, a multiple regression model with correlated predictors can indicate how well the entire bundle of predictors predicts the outcome variable, but it may not give valid results about any individual predictor, or about which predictors are redundant with respect to others.

### External validation/comparison with QOF prevalence

We used HSfE 2014 data to fit a logistic regression model for depression. As the local population estimates will most frequently be compared to QOF-registered prevalence, we carried out an external validation by examining the practice level depression prevalence estimates and QOF registered prevalence.

# Results

## Population and baseline characteristics

Figure 1: Hypertension population flowchart (Health Survey from England 2013+2014)

21,060 participants

16,872 participants

9,883 participants

4,188 participants under 16 years old

6,989 participants without three valid blood pressure measurements

The flowchart (Figure 1) above describes all the steps made in order to obtain our final population. After merging both surveys, there were a total of 21,060 respondents. From these, observations from individuals below the age of 16 (4,188 respondents) were dropped from the analysis. Additionally, all those from whom it was not possible to obtain three valid blood pressure measurements during the nurse visit (6,989 respondents) were also dropped. Our analysis was therefore performed in a sample of 9,883 participants.

Table 5 displays the frequency and percentage of each type of hypertensive group in our final population, stratified by age and sex. We can see that most of our final population is normotensive (66.07%), while the percentage of diagnosed hypertensive respondents (22.51%) is almost double the undiagnosed (11.42%).

Table 5: Frequency and weighted percentage of hypertension in the final population

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2013 | | 2014 | | Total | |
| Outcome | **Frequency** | **%** | **Frequency** | **%** | **Frequency** | **%** |
| Normotensive | 3,199 | 66.17% | 2,870 | 65.95% | 6,069 | 66.07% |
| Diagnosed hypertension | 1,378 | 22.63% | 1,212 | 22.37% | 2,590 | 22.51% |
| Undiagnosed hypertension | 637 | 11.19% | 587 | 11.68% | 1,224 | 11.42% |
| Total | **5,214** | **100.00%** | **4,669** | **100.00%** | **9,883** | **100.00%** |

Table 6 shows the distribution of respondents between the different hypertension categories commonly used in HSfE for the years 2003 to 2014 and stratified by sex.

Table 6: Blood pressure levels (%) for men and women during the period 2003-2014 as displayed in HSfE 2014 [90]

|  | 2003 | 2004d | 2005e | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | % | % | % | % | % | % | % | % | % | % | % | % |
| ALL MEN | | | | | | | | | | | | |
| Normotensive untreated | 68.3 | - | 66.7 | 68.8 | 68.9 | 68.3 | 68.0 | 68.5 | 68.9 | 69.1 | 68.6 | 67.6 |
| Hypertensive controlled | 5.4 | - | 7.4 | 6.8 | 8.0 | 8.3 | 8.3 | 10.3 | 10.6 | 9.4 | 9.2 | 9.9 |
| Hypertensive uncontrolled | 6.3 | - | 6.8 | 6.3 | 6.2 | 6.3 | 6.1 | 6.5 | 6.3 | 5.2 | 6.0 | 5.9 |
| Hypertensive untreated | 20.1 | - | 19.0 | 18.1 | 16.9 | 17.1 | 17.6 | 14.7 | 14.2 | 16.3 | 16.2 | 16.6 |
| All with high blood pressure | 31.7 | - | 33.3 | 31.2 | 31.1 | 31.7 | 32.0 | 31.5 | 31.1 | 30.9 | 31.4 | 32.4 |
| ALL WOMEN | | | | | | | | | | | | |
| Normotensive untreated | 70.5 | - | 72.7 | 72.2 | 71.0 | 71.4 | 73.1 | 71.0 | 72.0 | 73.2 | 74.0 | 73.1 |
| Hypertensive controlled | 6.0 | - | 7.9 | 7.8 | 8.4 | 9.2 | 7.2 | 10.9 | 10.2 | 9.4 | 10.3 | 9.9 |
| Hypertensive uncontrolled | 7.7 | - | 8.1 | 7.2 | 7.0 | 7.0 | 7.1 | 7.8 | 6.5 | 6.4 | 6.0 | 6.0 |
| Hypertensive untreated | 15.8 | - | 11.4 | 12.8 | 13.6 | 12.3 | 12.6 | 10.3 | 11.3 | 11.0 | 9.7 | 11.0 |
| All with high blood pressure | 29.5 | - | 27.3 | 27.8 | 29.0 | 28.6 | 26.9 | 29.0 | 28.0 | 26.8 | 26.0 | 26.9 |
| ALL ADULTS | | | | | | | | | | | | |
| 16-24 | | | | | | | | | | | | |
| Normotensive untreated | 95.8 | - | 95.3 | 96.2 | 95.1 | 95.2 | 96.5 | 96.1 | 94.9 | 96.6 | 96.7 | 97.1 |
| Hypertensive controlled | 0.1 | - | - | 0.2 | - | - | - | 0.4 | - | - | - | - |
| Hypertensive uncontrolled | - | - | - | 0.1 | - | - | - | - | 0.3 | - | - | - |
| Hypertensive untreated | 4.1 | - | 4.7 | 3.4 | 4.9 | 4.8 | 3.5 | 3.5 | 4.8 | 3.4 | 3.3 | 2.9 |
| All with high blood pressure | 4.2 | - | 4.7 | 3.8 | 4.9 | 4.8 | 3.5 | 3.9 | 5.1 | 3.4 | 3.3 | 2.9 |
| 25-34 | | | | | | | | | | | | |
| Normotensive untreated | 91.9 | - | 89.7 | 90.2 | 90.5 | 91.0 | 90.4 | 95.1 | 92.6 | 91.0 | 92.0 | 88.2 |
| Hypertensive controlled | 0.2 | - | 1.1 | 0.4 | 0.3 | 0.3 | - | 0.4 | 0.8 | 0.8 | 0.4 | 0.3 |
| Hypertensive uncontrolled | 0.3 | - | 0.3 | 0.2 | 0.2 | 0.3 | 0.6 | 0.3 | 0.1 | 0.2 | - | 0.2 |
| Hypertensive untreated | 7.6 | - | 8.8 | 9.2 | 9.1 | 8.4 | 8.9 | 4.2 | 6.4 | 8.1 | 7.6 | 11.4 |
| All with high blood pressure | 8.1 | - | 10.3 | 9.8 | 9.5 | 9.0 | 9.6 | 4.9 | 7.4 | 9.0 | 8.0 | 11.8 |
| 35-44 | | | | | | | | | | | | |
| Normotensive untreated | 84.6 | - | 82.4 | 87.2 | 86.2 | 84.3 | 85.7 | 82.9 | 84.9 | 88.0 | 86.0 | 85.0 |
| Hypertensive controlled | 1.5 | - | 1.6 | 1.6 | 1.6 | 2.2 | 2.7 | 2.9 | 2.8 | 2.1 | 2.2 | 1.6 |
| Hypertensive uncontrolled | 1.1 | - | 1.5 | 1.0 | 1.0 | 2.2 | 2.0 | 1.3 | 1.6 | 0.7 | 2.6 | 2.5 |
| Hypertensive untreated | 12.8 | - | 14.5 | 10.2 | 11.2 | 11.3 | 9.5 | 12.9 | 10.6 | 9.2 | 9.2 | 11.0 |
| All with high blood pressure | 15.4 | - | 17.6 | 12.8 | 13.8 | 15.7 | 14.3 | 17.1 | 15.1 | 12.0 | 14.0 | 15.0 |
| 45-54 | | | | | | | | | | | | |
| Normotensive untreated | 71.0 | - | 74.7 | 70.0 | 70.7 | 71.4 | 71.3 | 68.8 | 71.4 | 72.9 | 72.5 | 73.3 |
| Hypertensive controlled | 5.2 | - | 6.6 | 6.1 | 6.7 | 7.3 | 5.5 | 9.0 | 10.1 | 7.1 | 7.3 | 8.3 |
| Hypertensive uncontrolled | 3.6 | - | 5.0 | 4.9 | 5.1 | 4.4 | 4.3 | 4.5 | 3.7 | 3.4 | 4.0 | 2.7 |
| Hypertensive untreated | 20.2 | - | 13.7 | 19.1 | 17.5 | 17.0 | 18.9 | 17.7 | 14.9 | 16.6 | 16.2 | 15.8 |
| All with high blood pressure | 29.0 | - | 25.3 | 30.0 | 29.3 | 28.6 | 28.7 | 31.2 | 28.6 | 27.1 | 27.5 | 26.7 |
| 55-64 | | | | | | | | | | | | |
| Normotensive untreated | 51.5 | - | 55.2 | 56.7 | 53.5 | 53.3 | 52.8 | 51.6 | 56.0 | 55.0 | 56.8 | 54.8 |
| Hypertensive controlled | 11.2 | - | 14.8 | 11.7 | 14.2 | 15.3 | 11.1 | 18.4 | 14.8 | 14.3 | 16.3 | 17.8 |
| Hypertensive uncontrolled | 11.2 | - | 10.4 | 9.3 | 10.5 | 10.0 | 9.7 | 11.6 | 9.5 | 8.0 | 8.3 | 9.2 |
| Hypertensive untreated | 26.2 | - | 19.7 | 22.3 | 21.9 | 21.5 | 26.3 | 18.4 | 19.7 | 22.8 | 18.6 | 18.3 |
| All with high blood pressure | 48.5 | - | 44.8 | 43.3 | 46.5 | 46.7 | 47.2 | 48.4 | 44.0 | 45.0 | 43.2 | 45.2 |
| 65-74 | | | | | | | | | | | | |
| Normotensive untreated | 35.8 | - | 39.6 | 38.1 | 37.8 | 38.4 | 41.0 | 35.8 | 41.2 | 42.7 | 41.3 | 41.8 |
| Hypertensive controlled | 16.0 | - | 17.6 | 20.7 | 22.2 | 21.4 | 20.2 | 25.2 | 23.5 | 23.1 | 24.0 | 21.8 |
| Hypertensive uncontrolled | 18.4 | - | 17.4 | 16.4 | 17.0 | 18.1 | 16.5 | 20.0 | 15.1 | 14.6 | 15.5 | 15.1 |
| Hypertensive untreated | 29.9 | - | 25.5 | 24.8 | 23.0 | 22.1 | 22.3 | 19.0 | 20.3 | 19.6 | 19.1 | 21.3 |
| All with high blood pressure | 64.2 | - | 60.4 | 61.9 | 62.2 | 61.6 | 59.0 | 64.2 | 58.8 | 57.3 | 58.7 | 58.2 |
| 75+ | | | | | | | | | | | | |
| Normotensive untreated | 29.1 | - | 33.4 | 31.8 | 30.9 | 28.9 | 32.0 | 20.6 | 27.4 | 29.3 | 34.4 | 33.4 |
| Hypertensive controlled | 11.6 | - | 18.9 | 19.3 | 22.6 | 26.1 | 25.4 | 37.7 | 33.8 | 29.7 | 28.5 | 29.5 |
| Hypertensive uncontrolled | 25.7 | - | 24.0 | 25.2 | 21.9 | 21.9 | 22.9 | 28.7 | 23.7 | 22.5 | 18.6 | 19.2 |
| Hypertensive untreated | 33.6 | - | 23.7 | 23.7 | 24.6 | 23.1 | 19.7 | 13.1 | 15.1 | 18.5 | 18.5 | 17.9 |
| All with high blood pressure | 70.9 | - | 66.6 | 68.2 | 69.1 | 71.1 | 68.0 | 79.4 | 72.6 | 70.7 | 65.6 | 66.6 |

a In the 1998 report the systolic blood pressure (SBP) and diastolic blood pressure (DBP) thresholds for hypertension were changed from 160/95 to 140/90 mmHg, in accordance with the latest guidelines on hypertension management.1,2,3 From 2003, participants were placed in one of the treated categories if they were currently taking a drug prescribed for high blood pressure, whereas previously they had been described as treated if they were prescribed any drug which had the effect of lowering their blood pressure.  
*Refs*  
1 Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure and the National Blood Pressure Education Program Coordinating committee. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. Arch Intern Med 1997;157:2413-2446.  
2 1999 World Health Organisation - International Society of Hypertension Guidelines for Management of Hypertension. J Hypertens 1999;17:151-183.  
3 Ramsey LE, Williams B, Johnston GD, MacGregor GA, Poston L, Potter JF, Poulter NR, Russell G. British Hypertension Society guidelines for hypertension management 1999:Summary BMJ 1999;319:630-635.

b From 2003 the Dinamap monitor was replaced by the Omron. Omron measures are the preferred measure.

c Participants were classified into one of four groups as follows:

*Normotensive untreated*: SBP below 140mmHg and DBP below 90mmHg, not currently taking medication for blood pressure.

*Hypertensive controlled*: SBP below 140mmHg and DBP below 90mmHg, currently taking medication for blood pressure.

*Hypertensive uncontrolled*: SBP at or greater than 140mmHg and DBP at or greater than 90mmHg, currently taking medication for blood pressure.

*Hypertensive untreated*: SBP at or greater than 140mmHg and DBP at or greater than 90mmHg, not currently taking medication for blood pressure.

d Blood pressure was not measured in 2004.

e All adults from core and boost samples in 2005 were included in analysis of 65-74 and 75+ age groups but only the core sample was included in the overall total. It should therefore be noted that the 'All Men', 'All Women' and 'All adults' totals are not the sum of the individual age groups.

f Data from 2003 onwards are weighted for non-response.

## Statistical analyses

### Descriptive Statistics for the final population

Table 7 shows the descriptive statistics for demographics and main risk factors in our final population, stratified for the years 2013 and 2014 and in total. Both surveys display an almost identical distribution for each category, with one significant exception for fruit and legume consumption. In 2014 HSfE, all respondents in the sample reported consuming less than one piece a week. However, this is likely to be an artefact, as the overall number of respondents reporting more than that in the whole survey was minimal. Coding in this study was revised for mistakes, so this unlike feature comes directly from the survey.

Table 7: Distribution of risk and protective factor among the final population (N=9,883)

| Risk or Protective factor | | 2013 | | 2014 | | Total | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | **% (Weighted)** | **N** | **% (Weighted)** | **N** | **% (Weighted)** |
| Sex |  |  |  |  |  |  |  |
|  | Female | 2,910 | 52% | 2,612 | 51% | 5,522 | 51% |
|  | Male | 2,304 | 48% | 2,057 | 49% | 4,361 | 49% |
| Age |  |  |  |  |  |  |  |
|  | 16-24 | 395 | 13% | 356 | 14% | 751 | 14% |
|  | 25-34 | 677 | 16% | 569 | 16% | 1,246 | 16% |
|  | 35-44 | 821 | 17% | 762 | 16% | 1,583 | 17% |
|  | 45-54 | 960 | 18% | 872 | 18% | 1,832 | 18% |
|  | 55-64 | 864 | 14% | 756 | 14% | 1,620 | 14% |
|  | 65-74 | 873 | 12% | 798 | 12% | 1,671 | 12% |
|  | 75+ | 624 | 10% | 556 | 10% | 1,180 | 10% |
| Top qualification | |  |  |  |  |  |  |
|  | NVQ4/NVQ5/Degree or equiv. | 1,277 | 25% | 1,175 | 26% | 2,452 | 26% |
|  | Higher ed below degre | 584 | 11% | 557 | 11% | 1,141 | 11% |
|  | NVQ3/GCE A Level equiv. | 640 | 12% | 621 | 14% | 1,261 | 13% |
|  | NVQ2/GCE O Level equiv. | 989 | 18% | 845 | 17% | 1,834 | 18% |
|  | NVQ1/CSE other grade | 238 | 4% | 164 | 3% | 402 | 4% |
|  | Foreign/other | 91 | 1% | 63 | 1% | 154 | 1% |
|  | No qualification | 1,138 | 19% | 1,006 | 19% | 2,144 | 19% |
|  | FT Student | 248 | 9% | 233 | 9% | 481 | 9% |
| Employment category | |  |  |  |  |  |  |
|  | Higher managerial and professional | 626 | 12% | 566 | 12% | 1,192 | 12% |
|  | Lower managerial and professional | 1,165 | 22% | 1,083 | 22% | 2,248 | 22% |
|  | Intermediate occupations | 803 | 15% | 714 | 14% | 1,517 | 14% |
|  | Small employers and own account | 463 | 9% | 390 | 8% | 853 | 9% |
|  | Lower supervisory and technical | 360 | 7% | 302 | 6% | 662 | 7% |
|  | Semi-routine occupations | 907 | 18% | 784 | 17% | 1,691 | 17% |
|  | Routine occupations | 604 | 12% | 574 | 12% | 1,178 | 12% |
|  | Never worked and longterm unemployed | 89 | 2% | 82 | 2% | 171 | 2% |
|  | Other | 123 | 4% | 120 | 5% | 243 | 5% |
| Ethnicity | |  |  |  |  |  |  |
|  | White | 4,693 | 88% | 4,232 | 88% | 8,925 | 88% |
|  | Mixed | 52 | 1% | 48 | 1% | 100 | 1% |
|  | South Asian | 207 | 5% | 189 | 5% | 396 | 5% |
|  | African/Afro-Caribbean | 125 | 3% | 99 | 2% | 224 | 3% |
|  | Other background | 137 | 3% | 101 | 3% | 238 | 3% |
| IMD |  |  |  |  |  |  |  |
|  | 0.53->8.49 [least dep | 1,092 | 20% | 1,123 | 23% | 2,215 | 21% |
|  | 8.49->13.79 | 1,158 | 22% | 965 | 20% | 2,123 | 21% |
|  | 13.79->21.35 | 1,126 | 21% | 951 | 20% | 2,077 | 21% |
|  | 21.35->34.17 | 930 | 19% | 890 | 20% | 1,820 | 20% |
|  | 34.17->87.80 [most de | 908 | 18% | 740 | 17% | 1,648 | 17% |
| BMI |  |  |  |  |  |  |  |
|  | Under 18.5 | 109 | 3% | 83 | 2% | 192 | 3% |
|  | 18.5 and below 25 | 2,070 | 44% | 1,844 | 44% | 3,914 | 44% |
|  | 25 and below 30 | 1,745 | 35% | 1,646 | 35% | 3,391 | 35% |
|  | 30 and below 40 | 902 | 17% | 780 | 16% | 1,682 | 17% |
|  | Over 40 | 83 | 2% | 82 | 2% | 165 | 2% |
| Alcohol consumption (times per week) | | | |  |  |  |  |
|  | Less than once a week | 2,438 | 48% | 2,214 | 49% | 4,652 | 48% |
|  | Once or twice a week | 1,368 | 27% | 1,214 | 26% | 2,582 | 27% |
|  | Three or four times per week | 700 | 13% | 616 | 13% | 1,316 | 13% |
| Fruit consumption | |  |  |  |  |  |  |
|  | Less than two | 1,212 | 24% | 4,669 | 100% | 5,881 | 60% |
|  | Two to four | 2,532 | 49% | 0 | 0% | 2,532 | 26% |
|  | Five or more | 1,470 | 27% | 0 | 0% | 1,470 | 14% |
| Limiting long-lasting illness | | |  |  |  |  |  |
|  | No limiting longlasting illness | 2,842 | 59% | 2,590 | 59% | 5,432 | 59% |
|  | Non limiting longlasting illness | 983 | 17% | 869 | 17% | 1,852 | 17% |
|  | Limiting long-lasting illness | 1,387 | 24% | 1,204 | 23% | 2,591 | 24% |
| Physical activity (Tertile of moderate or vigorous intensive minutes of activity per week) | | | | | | | |
|  | Low | 1,718 | 35% | 1,465 | 34% | 3,183 | 34% |
|  | Medium | 1,479 | 33% | 1,349 | 35% | 2,828 | 34% |
|  | High | 1,433 | 32% | 1,217 | 31% | 2,650 | 32% |
| Smoking status | |  |  |  |  |  |  |
|  | Never smoked cigarettes | 2,574 | 51% | 2,485 | 54% | 5,059 | 53% |
|  | Used to smoke cigarettes occasionally | 272 | 5% | 256 | 6% | 528 | 5% |
|  | |  | | --- | | Used to smoked cigarettes  regularly | | 1,589 | 27% | 1,276 | 25% | 2,865 | 26% |
|  | Current cigarette smoker | 767 | 17% | 642 | 15% | 1,409 | 16% |

### Missing values

Table 8 shows the number and percentage of missing values in total and stratified per year. The biggest percentage of missing values was found in the variable self-reported physical activity, followed by self-reported BMI. With these two exceptions, all other variables were missing less than 1.5% of their observations. Missing values from the variable origin where recoded to be considered within the category “White”, while physical activity and smoking status were finally excluded from both models. The remaining missing values where inferred for analysis using multiple imputation.

Table 8: Missing values in the final population.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Risk or Protective factor | | 2013 | | | 2014 | | Total | |
|  | | **Frequency** | | **%\*** | **Frequency** | **%\*** | **Frequency** | **%\*** |
| Top qualification | | 9 | 0.20% | | 5 | 0.10% | 14 | 0.15% |
| Employment category | | 74 | 1.58% | | 54 | 1.31% | 128 | 1.47% |
| BMI | | 305 | 6.28% | | 234 | 5.29% | 539 | 5.81% |
| Alcohol consumption | | 15 | 0.49% | | 13 | 0.51% | 28 | 0.50% |
| Limiting long-lasting illness | | 2 | 0.03% | | 6 | 0.16% | 8 | 0.09% |
| Physical activity | | 584 | 11.20% | | 638 | 13.22% | 1222 | 12.15% |
| Smoking status | | 12 | 0.42% | | 10 | 0.32% | 22 | 0.38% |
| Ethnic Origin | | 34 | 0.02% | | 39 | 0.05% | 73 | 0.04% |
|  | \*Weighted |  |  | |  |  |  |  |

### Univariable analyses

Table 9 shows the results of univariablee models for individual risk factors and diagnosed hypertension. Considering the variables unadjusted, only fruit consumption seems likely to not be correlated with the odds of having diagnosed hypertension. As observed in the literature , older age, lower qualification, lower job category, higher BMI, existence of a long-lasting disease and less physical exercise are likely to be associated with greater odds of having diagnosed hypertension.

Table 9: Univariable analysis of risk factors for diagnosed hypertension

| Predictor variable | Odds ratio | P>z | [95% Conf. Interval] |
| --- | --- | --- | --- |
| Sex |  |  |  |
| Male | 1.10 | 0.04 | (1.00-1.20) |
| Age |  |  |  |
| 25-34 | 4.27 | 0.00 | (2.04-8.95) |
| 35-44 | 9.66 | 0.00 | (4.66-20.00) |
| 45-54 | 19.17 | 0.00 | (9.46-38.87) |
| 55-64 | 40.16 | 0.00 | (19.98-80.75) |
| 65-74 | 71.15 | 0.00 | (35.33-143.29) |
| 75+ | 104.93 | 0.00 | (52.63-209.22) |
| Top qualification |  |  |  |
| Higher ed below degree | 1.63 | 0.00 | (1.36-1.96) |
| NVQ3/GCE A Level equiv. | 1.09 | 0.37 | (0.90-1.31) |
| NVQ2/GCE O Level equiv. | 1.53 | 0.00 | (1.30-1.80) |
| NVQ1/CSE other grade | 1.67 | 0.00 | (1.29-2.16) |
| Foreign/other | 4.28 | 0.00 | (3.02-6.06) |
| No qualification | 4.20 | 0.00 | (3.62-4.87) |
| FT Student | 0.10 | 0.00 | (0.05-0.19) |
| Employment category |  |  |  |
| Lower managerial and professional | 1.27 | 0.01 | (1.06-1.51) |
| Intermediate occupations | 1.37 | 0.00 | (1.13-1.65) |
| Small employers and own account | 1.28 | 0.03 | (1.02-1.60) |
| Lower supervisory and technical | 1.92 | 0.00 | (1.54-2.41) |
| Semi-routine occupations | 1.49 | 0.00 | (1.24-1.79) |
| Routine occupations | 1.92 | 0.00 | (1.57-2.34) |
| Never worked and longterm unemployed | 0.85 | 0.45 | (0.56-1.30) |
| Other | 0.06 | 0.00 | (0.01-0.24) |
| Ethnicity |  |  |  |
| Mixed | 0.55 | 0.03 | (0.33-0.93) |
| South Asian | 0.77 | 0.04 | (0.60-0.98) |
| African/Afro-Caribbea | 1.43 | 0.04 | (1.02-1.99) |
| Other background | 0.43 | 0.00 | (0.30-0.61) |
| IMD |  |  |  |
| 8.49->13.79 | 1.09 | 0.26 | (0.94-1.27) |
| 13.79->21.35 | 1.16 | 0.07 | (0.99-1.35) |
| 21.35->34.17 | 0.98 | 0.81 | (0.83-1.15) |
| 34.17->87.80 | 1.05 | 0.56 | (0.89-1.25) |
| BMI |  |  |  |
| 18.5 and below 25 | 1.35 | 0.16 | (0.89-2.05) |
| 25 and below 30 | 2.88 | 0.00 | (1.91-4.36) |
| 30 and below 40 | 4.85 | 0.00 | (3.20-7.35) |
| Over 40 | 6.07 | 0.00 | (3.56-10.36) |
| Alcohol consumption (times per week) |  |  |  |
| Once or twice a week | 0.80 | 0.00 | (0.70-0.91) |
| Three or four times p | 0.88 | 0.12 | (0.75-1.03) |
| Five or more | 1.44 | 0.00 | (1.24-1.66) |
| Fruit consumption |  |  |  |
| Once or twice a week | 1.05 | 0.37 | (0.94-1.18) |
| Two to four | 0.88 | 0.12 | (0.84-1.13) |
| Limiting long-lasting illness |  |  |  |
| Non limiting long-lasting illness | 4.69 | 0.00 | (4.10-5.35) |
| No limiting long-lasting illness | 5.65 | 0.00 | (5.00-6.39) |
| Physical activity (Tertile of moderate or vigorous intensive minutes of activity per week) |  |  |  |
| Medium | 0.47 | 0.00 | (0.41-0.53) |
| High | 0.47 | 0.00 | (0.41-0.53) |
| Smoking status |  |  |  |
| Used to smoke cigarettes occasionally | 0.91 | 0.43 | (0.72-1.15) |
| Used to smoke cigarettes regularly | 1.74 | 0.00 | (1.56-1.94) |
| Current cigarette smoker | 0.76 | 0.00 | (0.64-0.91) |

Table 10 shows the results of univariate models for individual risk factors and undiagnosed hypertension. Considering the variables unadjusted, job category, deprivation level, alcohol consumption, physical activity and having a long-lasting disease would not seem to be correlated with the odds of having undiagnosed hypertension. However, it must be bear in mind that this is an unadjusted analysis when comparing it with the results obtained from the multivariate logistic model.

Table 10: Univariable analysis of risk factors for undiagnosed hypertension.

| Predictor variable | Odds ratio | P>z | [95% Conf. Interval] |
| --- | --- | --- | --- |
| Sex |  |  |  |
| Male | 1.57 | 0.00 | (1.38-1.78) |
| Age |  |  |  |
| 25-34 | 2.97 | 0.00 | (1.70-5.19) |
| 35-44 | 2.96 | 0.00 | (1.79-4.90) |
| 45-54 | 4.89 | 0.00 | (2.94-8.12) |
| 55-64 | 6.32 | 0.00 | (3.87-10.33) |
| 65-74 | 7.06 | 0.00 | (4.30-11.58) |
| 75+ | 7.93 | 0.00 | (4.74-13.26) |
| Top qualification |  |  |  |
| Higher ed below degree | 1.42 | 0.00 | (1.13-1.78) |
| NVQ3/GCE A Level equiv. | 1.14 | 0.29 | (0.89-1.45) |
| NVQ2/GCE O Level equiv. | 1.37 | 0.00 | (1.11-1.69) |
| NVQ1/CSE other grade | 1.34 | 0.09 | (0.96-1.87) |
| Foreign/other | 2.21 | 0.00 | (1.42-3.43) |
| No qualification | 1.58 | 0.00 | (1.30-1.91) |
| FT Student | 0.27 | 0.00 | (0.15-0.46) |
| Employment category |  |  |  |
| Lower managerial and professional | 0.81 | 0.08 | (0.64-1.03) |
| Intermediate occupations | 0.91 | 0.46 | (0.70-1.17) |
| Small employers and own account | 1.31 | 0.06 | (0.99-1.73) |
| Lower supervisory and technical | 1.28 | 0.11 | (0.95-1.72) |
| Semi-routine occupations | 1.00 | 0.99 | (0.78-1.29) |
| Routine occupations | 0.99 | 0.95 | (0.76-1.30) |
| Never worked and longterm unemployed | 0.73 | 0.24 | (0.42-1.25) |
| Other | 0.16 | 0.00 | (0.07-0.38) |
| Ethnicity |  |  |  |
| Mixed | 0.58 | 0.11 | (0.29-1.13) |
| South Asian | 0.81 | 0.28 | (0.55-1.19) |
| African/Afro-Caribbean | 0.59 | 0.03 | (0.36-0.95) |
| Other background | 0.64 | 0.05 | (0.41-0.99) |
| IMD |  |  |  |
| 8.49->13.79 | 1.13 | 0.21 | (0.93-1.37) |
| 13.79->21.35 | 1.17 | 0.12 | (0.96-1.42) |
| 21.35->34.17 | 0.94 | 0.59 | (0.77-1.16) |
| 34.17->87.80 | 0.97 | 0.75 | (0.77-1.20) |
| BMI |  |  |  |
| 18.5 and below 25 | 1.90 | 0.06 | (0.98-3.68) |
| 25 and below 30 | 3.21 | 0.00 | (1.66-6.23) |
| 30 and below 40 | 3.27 | 0.00 | (1.66-6.45) |
| Over 40 | 4.19 | 0.00 | (1.85-9.50) |
| Alcohol consumption (times per week) |  |  |  |
| Once or twice a week | 1.01 | 0.87 | (0.85-1.20) |
| Three or four times p | 1.00 | 0.99 | (0.82-1.23) |
| Five or more | 2.09 | 0.41 | (1.74-2.50) |
| Fruit consumption |  |  |  |
| Once or twice a week | 0.95 | 0.00 | (0.00-0.00) |
| Two to four |  | 1.11 | (0.00-0.00) |
| Limiting long-lasting illness |  |  |  |
| Non limiting long-lasting illness | 1.00 | 0.98 | (0.83-1.20) |
| No limiting long-lasting illness | 1.01 | 0.90 | (0.87-1.17) |
| Physical activity (Tertile of moderate or vigorous intensive minutes of activity per week) |  |  |  |
| Medium | 0.88 | 0.14 | (0.75-1.04) |
| High | 0.90 | 0.19 | (0.76-1.06) |
| Smoking status |  |  |  |
| Used to smoke cigarettes occasionally | 0.88 | 0.46 | (0.64-1.23) |
| Used to smoke cigarettes regularly | 1.30 | 0.00 | (1.13-1.49) |
| Current cigarette smoker | 1.04 | 0.69 | (0.85-1.27) |

### Multivariable analyses

Table 11 shows the odds ratios of different risk factors after fitting a multivariate logistic regression model for diagnosed hypertension. The values observed are those obtained are adjusted for all variables in the model and correspond to Model D0, defined before introducing interactions. After using a forward stepwise approach and Wald tests to select those variables observed to be significant, only the variables for sex, age, education, ethnic origin, employment category, BMI, long-lasting illness were retained in the model. The variable for the index of multiple deprivation was not found significant for diagnosed neither undiagnosed hypertension. Nevertheless, given the importance of the variable, it was decided to retain it in both multivariate models to adjust the results for socioeconomic differences.

Table 11: Multivariate analysis for diagnosed hypertension 2013-2014, complete model without interactions

| Predictor variable | | Odds ratio | P>z | [95% Conf. Interval] |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Sex |  |  |  |  |
| Male | | 1.27 | 0.00 | (1.12-1.44) |
|  |  |  |  |  |
| Age |  |  |  |  |
|  | 25-34 | 2.91 | 0.01 | (1.30-6.48) |
|  | 35-44 | 5.63 | 0.00 | (2.52-12.59) |
|  | 45-54 | 10.32 | 0.00 | (4.70-22.67) |
|  | 55-64 | 20.44 | 0.00 | (9.33-44.77) |
|  | 65-74 | 34.00 | 0.00 | (15.38-75.17) |
|  | 75+ | 47.16 | 0.00 | (21.63-102.79) |
|  |  |  |  |  |
| Top qualification | |  |  |  |
|  | Higher ed below degre | 0.99 | 0.93 | (0.79-1.23) |
|  | NVQ3/GCE A Level equiv. | 1.00 | 0.99 | (0.80-1.25) |
|  | NVQ2/GCE O Level equiv. | 0.98 | 0.83 | (0.79-1.20) |
|  | NVQ1/CSE other grade | 0.91 | 0.58 | (0.67-1.25) |
|  | Foreign/other | 1.19 | 0.41 | (0.79-1.81) |
|  | No qualification | 1.33 | 0.01 | (1.08-1.64) |
|  | FT Student | 0.56 | 0.09 | (0.29-1.10) |
|  |  |  |  |  |
| Ethnicity |  |  |  |  |
|  | Mixed | 1.06 | 0.86 | (0.56-2.01) |
|  | South Asian | 1.70 | 0.00 | (1.29-2.24) |
|  | African/Afro-Caribbean | 2.48 | 0.00 | (1.69-3.63) |
|  | Other background | 1.20 | 0.38 | (0.80-1.82) |
|  |  |  |  |  |
| Employment category | | |  |  |
|  | Lower managerial and professional | 1.22 | 0.06 | (0.99-1.50) |
|  | Intermediate occupations | 1.18 | 0.18 | (0.93-1.50) |
|  | Small employers and own account | 0.83 | 0.17 | (0.63-1.08) |
|  | Lower supervisory and technical | 1.28 | 0.10 | (0.96-1.72) |
|  | Semi-routine occupations | 1.16 | 0.23 | (0.91-1.48) |
|  | Routine occupations | 1.32 | 0.04 | (1.02-1.73) |
| Never worked and longterm unemployed | | 0.88 | 0.62 | (0.52-1.47) |
|  | Other | 1.11 | 0.90 | (0.22-5.54) |
|  |  |  |  |  |
| IMD |  |  |  |  |
|  | 8.49->13.79 | 1.08 | 0.32 | (0.93-1.27) |
|  | 13.79->21.35 | 1.14 | 0.15 | (0.96-1.36) |
|  | 21.35->34.17 | 1.12 | 0.23 | (0.93-1.35) |
|  | 34.17->87.80 | 1.21 | 0.05 | (1.00-1.48) |
|  |  |  |  |  |
| BMI |  |  |  |  |
|  | 18.5 to below 25 | 0.92 | 0.73 | (0.56-1.49) |
|  | 25 to below 30 | 1.52 | 0.09 | (0.94-2.46) |
|  | 30 to below 40 | 2.58 | 0.00 | (1.58-4.22) |
|  | Over 40 | 3.31 | 0.00 | (1.71-6.43) |
|  |  |  |  |  |
| Limiting long-lasting illness | | |  |  |
|  | Non limiting long-lasting illness | 2.81 | 0.00 | (2.44-3.22) |
|  | No limiting long-lasting illness | 2.70 | 0.00 | (2.36-3.10) |

Table 12 shows the odds ratios of different risk factors after fitting a multivariate logistic regression model for undiagnosed hypertension. The values observed are those obtained are adjusted for all variables in the model and correspond to Model Und0, defined before introducing interactions. In this case, a smaller number of variables were retained after fitting the model. The model shares the variables sex, age, BMI and limiting longlasting illness with the diagnosed hypertension model, and likewise, the variable for the index of multiple deprivation was retained to adjust the model for socioeconomic differences. In addition, alcohol consumption, which was not found significant for diagnosed hypertension, is included in this model.

Table 12: Multivariate analysis for undiagnosed hypertension 2013-2014, complete model

| Predictor variable | | Odds ratio | P>z | [95% Conf. Interval] |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Sex |  |  |  |  |
| Male | | 1.55 | 0.00 | (1.36-1.77) |
|  |  |  |  |  |
| Age |  |  |  |  |
|  | 25-34 | 2.70 | 0.00 | (1.56-4.67) |
|  | 35-44 | 2.67 | 0.00 | (1.63-4.38) |
|  | 45-54 | 4.53 | 0.00 | (2.75-7.46) |
|  | 55-64 | 6.04 | 0.00 | (3.72-9.79) |
|  | 65-74 | 7.06 | 0.00 | (4.34-11.50) |
|  | 75+ | 8.88 | 0.00 | (5.35-14.76) |
|  |  |  |  |  |
| BMI |  |  |  |  |
|  | 18.5 to below 25 | 1.29 | 0.44 | (0.67-2.50) |
|  | 25 to below 30 | 1.79 | 0.08 | (0.93-3.47) |
|  | 30 to below 40 | 1.92 | 0.06 | (0.98-3.77) |
|  | Over 40 | 3.07 | 0.01 | (1.33-7.07) |
|  |  |  |  |  |
| Alcohol consumption (times per week) | | | |  |
|  | Once or twice a week | 0.96 | 0.64 | (0.80-1.15) |
|  | Three or four times p | 0.86 | 0.15 | (0.70-1.06) |
|  | Five or more | 1.44 | 0.00 | (1.18-1.75) |
|  |  |  |  |  |
| Limiting long-lasting illness | | |  |  |
|  | Non limiting long-lasting illness | 0.67 | 0.00 | (0.55-0.81) |
|  | No limiting long-lasting illness | 0.61 | 0.00 | (0.52-0.73) |
|  |  |  |  |  |
| IMD |  |  |  |  |
|  | 8.49->13.79 | 1.12 | 0.24 | (0.93-1.35) |
|  | 13.79->21.35 | 1.20 | 0.07 | (0.99-1.46) |
|  | 21.35->34.17 | 1.12 | 0.28 | (0.91-1.39) |
|  | 34.17->87.80 | 1.21 | 0.09 | (0.97-1.52) |

## Internal validation

### Tests for interactions and collinearity

Interactions between the different variables of interest were tested in our “Model 0”, which included all the variables found significant in the multivariable analysis. Interactions were tested individually using Wald tests to determine whether they were significant or not. Interactions found significant were included to the model 0, creating a set of alternative models for which performance was compared using ROC curves. After testing a possible interaction, Wald tests were also conducted to confirm whether the original variables should be kept or not in the model. This process was performed for both models, and the most significant interactions for diagnosed (Table 13) and undiagnosed hypertension (Table 14) are shown below. In both tables only significant p-values from the interaction between two variables are shown, however, the value of the Wald test refers to the entire interaction variable.

Table 13: tests for interactions/effect modification in multivariable model 0 for diagnosed hypertension

| Interaction | | Interaction status/p-value | Wald test (Prob>F) |
| --- | --- | --- | --- |
| Sex | **Age** |  |  |
| Male | 35-44 | 0.007 | 0.0006 |
| Male | 45-54 | <0.001 |
| Sex | **Employment category** |  |  |
| Male | Lower managerial and professional occupations | 0.001 | 0.0126 |
| Male | Lower supervisory and technical occupations | 0.006 |
| Male | Semi-routine occupations | 0.001 |
| Male | Routine occupations | 0.002 |
| Sex | **IMD** |  |  |
| Male | 34.17->87.80 [most deprived] | 0.003 | 0.0079 |
| Sex | **BMI** |  |  |
| Male | 18.5 to below 25 | 0.014 | 0.0024 |
| Male | 25 to below 30 | 0.005 |
| Male | 30 to below 40 | 0.002 |
| Male | Over 40 | 0.001 |
| Age | **Limiting long-lasting illness** |  |  |
| 75+ | Limiting long-lasting illness | 0.037 | 0.0009 |
| Ethnicity | **Limiting long-lasting illness** |  |  |
| Mixed | Limiting long-lasting illness | 0.001 | 0.0024 |
| South Asian | Limiting long-lasting illness | 0.044 |
| IMD | **Limiting long-lasting illness** |  |  |
| 13.79->21.35 | No limiting longlasting illness | 0.047 | 0.0457 |
| 21.35->34.17 | Limiting longlasting illness | 0.020 |
| 34.17->87.80 | Limiting longlasting illness | 0.006 |

Table 14: tests for interactions/effect modification in undiagnosed hypertension multivariable model 4

| Interaction | | Interaction status/p-value | Wald test (Prob>F) |
| --- | --- | --- | --- |
| Sex | **Age** |  |  |
| Male | 35-44 | 0.020 | <0.0001 |
| Male | 45-54 | 0.012 |
| Male | 55-64 | 0.005 |
| Male | 65-74 | 0.001 |
| Male | 75+ | 0.001 |
| Sex | **Limiting long-lasting illness** |  |  |
| Male | Non limiting longlasting illness | <0.001 | <0.0001 |
| Male | Limiting longlasting illness | <0.001 |

Variance inflation factor (VIF) was used to check for collinearity between independent variables within the logistic model. If the VIF value is greater than 10 it suggests collinearity is present. Table 15 shows the VIF values among the risk factor variables within the diagnosed hypertension model. **Error! Reference source not found.** show the same for undiagnosed hypertension. There is no evidence of collinearity in any of the models.

Table 15: VIF values in the diagnosed hypertension complete model

|  |  |  |
| --- | --- | --- |
| Variable | VIF | Tolerance (1/VIF ) |
| Sex | 1.01 | 0.9869 |
| Age | 1.37 | 0.7312 |
| Top qualification | 1.32 | 0.7550 |
| Ethnicity | 1.05 | 0.9482 |
| Employment category | 1.33 | 0.7501 |
| IMD | 1.11 | 0.9042 |
| BMI | 1.05 | 0.9492 |
| Limiting long-lasting illness | 1.17 | 0.8551 |
|  |  |  |
| Mean VIF | 1.18 |  |

Table 16: VIF values in the undiagnosed hypertension complete model

|  |  |  |
| --- | --- | --- |
| Variable | VIF | 1/VIF |
| Sex | 1.04 | 0.9638 |
| Age | 1.19 | 0.8394 |
| BMI | 1.04 | 0.9585 |
| IMD | 1.09 | 0.9149 |
| Alcohol consumption | 1.17 | 0.8569 |
| Limiting long-lasting illness | 1.06 | 0.9447 |
|  |  |  |
| Mean VIF | 1.1 |  |

## Discrimination

### ROC curves

We next evaluated model performance using the ***mi predict*** command and receiver operating characteristics (ROC) curves for the various models. The best ROC curve which predicts data perfectly will touch the top-left corner of the plot (area 1.0), and the larger the area under the ROC curve the better the prediction. An area of 0.5 signifies a prediction no better than chance. However, as noted in the Methods, the choice of variables and hence the final model also depends on the availability of local data, so the final local model will not predict as well as the optimal/”gold standard” model i.e. shows the ROC curves for the final/”gold standard” model and the local model. The graphs for the all the different models’ ROC curves are shown in Annex 1: ROC curves.

Table 17: comparison of different models for diagnosed hypertension

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model description | Model | ROC area | SE | 95% CI |
| Stepwise forward model | DH0 | 0.8241 | 0.0044 | [0.8154-0.8328] |
| Stepwise forward model with sex-age interaction | DH1 | 0.8250 | 0.0044 | [0.8164-0.8337] |
| Stepwise forward model with sex-employment category interaction | DH2 | 0.8249 | 0.0044 | [0.8163-0.8336] |
| Stepwise forward model with sex-IMD interaction | DH3 | 0.8246 | 0.0044 | [0.8159-0.8333] |
| Stepwise forward model with sex-BMI interaction | DH4 | 0.8250 | 0.0044 | [0.8164-0.8337] |
| Stepwise forward model with age-limiting long-lasting illness interaction | DH5 | 0.8254 | 0.0044 | [0.8167-0.8340] |
| Stepwise forward model with ethnicity-limiting long-lasting illness interaction | DH6 | 0.8250 | 0.0044 | [0.8163-0.8337] |
| Stepwise forward model with IMD- limiting long-lasting illness interaction | DH7 | 0.8250 | 0.0044 | [0.8163-0.8337] |
| Stepwise forward model with sex-BMI and age- limiting long-lasting illness interactions | DH8 | 0.8263 | 0.0044 | [0.8177-0.8350] |

Figure 2: ROC curve for model DH8 (complete model for diagnosed hypertension)



Table 18: Comparison of different models for undiagnosed hypertension

| Model description | Model | ROC area | SE | 95% CI |
| --- | --- | --- | --- | --- |
| Stepwise forward model | UDH0 | 0.7041 | 0.0055 | [0.6934-0.7149] |
| Stepwise forward model with sex-age interaction | UDH1 | 0.6984 | 0.0055 | [0.6877-0.7091] |
| Stepwise forward model with sex-limiting long-lasting illness interaction | UDH2 | 0.7112 | 0.0054 | [0.7007-0.7217] |

### Predicted probabilities of having hypertension

We used the HSfE automatic stepwise forward and backward models combined to predict the probability of individual having hypertension in HSfE data set. We used box plots to show the predicted probability of people having hypertension among the non-hypertension and hypertension groups. Since we have a binary response model, we can choose a cut-off point on the predicted probability to separate the predicted hypertension cases (with higher predicted probability) from the predicted non-hypertension cases (with lower predicted probability).

### Sensitivity and specificity analysis

The sensitivity/specificity versus probability cut-off plots showed us the corresponding sensitivity and specificity in each possible probability cut-off point. Higher sensitivity would usually yield low specificity and vice versa, the rule of thumb is to choose a cut-off probability to maximize both. We chose the cut-off probability where sensitivity and specificity lines cross.

Figure 3: ROC curve for model UDH 2 (complete model for undiagnosed hypertension)



## Local estimates

Several changes were made in the local model for both diagnosed and undiagnosed hypertension in order to adjust it to the local data available. These included recoding some of the variables, dropping the alcohol consumption variable and using versions of the models not including interactions (Models DH0 and UDH0). Table 19 includes a summary of differences between the models.

In the first place, full time students are not considered as a separate category within the variable topqual, but are included within the others. Similarly, our variable for BMI switches from five to four categories, collapsing categories “30 to 40” and “40+” into “30+”. Similarly, local data for limiting illness was obtained in three categories depending how much was day-to-day activity was affected: not limited, limited a little and limited a lot. However, HSfE used three different categories: no limiting long-lasting illness, non-limiting long-lasting illness and limiting long-lasting illness. In order to maximize the comparability of these two sources, data was collapsed into two categories in each case. For local data, individuals whose day-to-day activities were limited were grouped together. Similarly, individuals with no limiting long-lasting disease and non-limiting long-lasting disease were also grouped together. In the complete model, we observed that individuals with non-limiting long-lasting diseases had increased odds of suffering both diagnosed and undiagnosed hypertension. After merging that group along with no limiting long-lasting illness, we are probably diluting the effect of the variable. Moreover, it is possible that individuals whose day-today activities are limited only a little don’t have the same odds of hypertension that those whose activities are limited a lot.

For undiagnosed hypertension, the original alcohol variable was excluded from the local model, due to incompatibility between the format of the data collected locally by PHE, available at the Local Alcohol Profiles for England (LAPE), and HSfE. Whilst in the former categories are made depending on the number of units consumed by an individual in a week period, HSfE questioned participants only about the frequency of their alcohol consumption during the last 12 months, irrespective of the number of drinks neither type of drink at each occassion. In addition, the number of units required to belong to one category or another in LAPE was gender-dependent, requiring females a smaller amount of units to enter a category. This distinction is not accounted for in HSfE. As a result, the alcohol variable was recoded in two categories, depending on whether they currently consumed alcohol or not. However, after testing its fitness within the model, it was decided to drop the variable as it was not significant.

Another significant difference between the complete and the local models can be find in the lack of interactions in the latter. Although we found that the inclusion of interaction variables could improve the performance of the model to some extent (please see Table 17 for diagnosed hypertension and Table 18 for undiagnosed hypertension), the limited amount of time available was not enough to adjust the inverse probability weights code to provide local estimates considering interactions.

Finally, when combining data from LSOAs or MSOAs to the practice level, we faced a problem as we lacked information from some of these. In order to be able to provide information for these practices, we exclude missing LSOAs/MSOAs from the calculations for each practice. As a result, we were able to provide data on all practices with LSOAs/MSOAs with available data, although the weight of the characteristics of the remaining units in each practice is higher than in a situation in which all information was available. Considering that one practice can include several MSOAs and vice versa, and the fact that the LSOAs/MSOAs’ areas from a same practice are very close from each other, the effect of our exclusion will presumably be small in most cases.

Table 19: Differences between the complete and the local model

|  |  |
| --- | --- |
| Modifications in the local model | Reasoning |
|  |  |
| The education variable was recoded, removing full-time students a separate category. | Required to make HSfE data comparable with local data. |
| Two categories from the variable “limiting long-lasting illness” were collapsed. | Required to make HSfE data comparable with local data. |
| Two categories from the variable “self-reportedBMI” were collapsed. | Required to make HSfE data comparable with local data. |
| The alcohol consumption variable was dropped from the undiagnosed hypertension model | Local data available was not comparable with that from the HSfE. |
| Interactions included in the final models were dropped. | Time limitation. |

After the changes mentioned above, performance of the local models varied in comparison with the complete models (DH8 and UDH2) but also with models DH0 and UDH0 (Table 20). Differences between the latter and local models are due to variable recoding and, in the case of undiagnosed hypertension, the dropping of the alcohol consumption variable. Interactions included in the final models were also dropped because there was insufficient time available to modify the inverse probability weights method Stata do-files.

Table 20: Comparison of the complete and local models for diagnosed and undiagnosed hypertension

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model description | Model | ROC area | SE | 95% CI |
| Initial model for diagnosed hypertension | DH0 | 0.8241 | 0.0044 | [0.8154-0.8328] |
| Complete model for diagnosed hypertension | DH8 | 0.8263 | 0.0044 | [0.8177-0.8350] |
| Local model for diagnosed hypertension | - | 0.8121 | 0.0046 | [0.8031-0.8211] |
| Initial model for undiagnosed hypertension | UDH0 | 0.7041 | 0.0055 | [0.6934-0.7149] |
| Complete model for undiagnosed hypertension | UDH2 | 0.7112 | 0.0054 | [0.7007-0.7217] |
| Local model for undiagnosed hypertension | - | 0.6535 | 0.0075 | [0.6389-0.6681] |

Local estimates and CIs were calculated by the inverse probability weights method.[91] Table 21 shows descriptive statistics for the CCG and practice level local estimates. These showed a mean prevalence of 21.30%. The average difference between estimated and QOF prevalence of 6.85%.

Table 21: summary statistics for CCG and practice level estimates

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CCG level | | Practice level | |
|  | **Percentiles** | **Smallest/Largest** |  |  |
| 1% | 14.39 | 13.38 | 9.95% | 4.29% |
| 5% | 15.94 | 13.63 | 14.75% | 7.78% |
| 10% | 17.26 | 14.39 | 16.42% | 9.33% |
| 25% | 19.50 | 14.82 | 18.97% | 11.88% |
| 50% | 21.30 |  | 21.23% | 14.15% |
| 75% | 22.62 | 25.42 | 23.10% | 16.31% |
| 90% | 23.81 | 25.51 | 24.74% | 18.27% |
| 95% | 24.48 | 26.33 | 25.74% | 19.65% |
| 99% | 25.51 | 26.77 | 28.33% | 22.69% |

### Internal validation of local estimates

Table 24 compares practice-level and aggregate numbers and prevalence derived from the local practice-level estimates with corresponding QOF register data for England Regions. The England values for diagnosed and undiagnosed hypertension from HSfE 2013+2014 are 22.51% and 11.42% respectively within the over 16 population. From QOF, the highest prevalence rates are for hypertension (13.8 per cent), obesity (9.0 per cent) and depression (7.3 per cent) The QOF hypertension register .

The bottom row shows the percentage difference between the local estimates and QOF registers. The largest differences appear to be in the South of England. The overall percentage difference between the local estimates and QOF registers is 12%. In general the local estimates are higher than the registered prevalence, as we would expect given the model we developed. The prevalence of GP-registered plus probable/possible cases in our CPRD dataset is about 20% higher than GP-registered prevalence alone, and the average prevalence in our local estimates is 15% higher than aggregated GP registers.

### External validation/comparison with QOF prevalence

The funding for the project does not include an in-depth external validation. For example, this could be carried out by obtaining an extract from a similar dataset e.g. applying the HSfE prevalence models’ equations to Whitehall II data. However another useful external data source is the Quality & Outcomes Framework (QOF) GP-diagnosed hypertension prevalence. This can obviously be compared with diagnosed hypertension prevalence from the model, taking into account that the HSfE definition was derived from the number of patients that reported being told by a nurse or doctor that they had high BP. Table 22 compares aggregate local estimates with aggregate QOF registers for England Regions.

Table 22: percentile differences between CCG-level model-estimated and QOF prevalence of diagnosed hypertension

| Percent | Percentile | (95% | CI) |
| --- | --- | --- | --- |
| 0 | 4.7 | (4.7-4.7) | |
| 25 | 6.4 | (6.3-6.6) | |
| 50 | 7.0 | (6.9-7.2) | |
| 75 | 7.6 | (7.4-7.7) | |
| 100 | 9.2 | (9.2-9.2) | |

It is sometimes possible to compare Regional breakdowns as a form of external validation. However this is very difficult with HSfE and QOF data because NHS Regions and old Government Office Regions are no longer co-terminous. We therefore carried out a disagreement analysis between model-estimated and QOF prevalence (%) of diagnosed hypertension in CCGs and practices. We estimated three principal components of disagreement (discordance as measured by Kendall's tau-a, bias as measured by percentile differences, and scale discrepancy as measured by Kendall’s tau-a between the mean of the two prevalences and the difference between the two prevalences). Confidence intervals were calculated for Kendall’s tau-a using the methods of Newson (2006a) [92], and for percentile differences using the methods of Newson (2006b) [93].

The Kendall's tau-a between model-estimated and QOF prevalence of diagnosed hypertension for 209 CCGs was 0.775 (95% Cis 0.736-0.809; *P*=1.4x10-57. Therefore, in our sample of CCGs, a random pair of CCGs was 77.5 percent more likely to be concordant (the CCG with the higher QOF prevalence also having the higher model-estimated prevalence) than to be discordant (the CCG with the higher QOF prevalence having the lower model-estimated prevalence). Figure 4 shows a Bland-Altman plot for CCG-level model-estimated and QOF prevalence of diagnosed hypertension.

Figure 4: Bland-Altman plot for CCG-level model-estimated and QOF prevalence of diagnosed hypertension

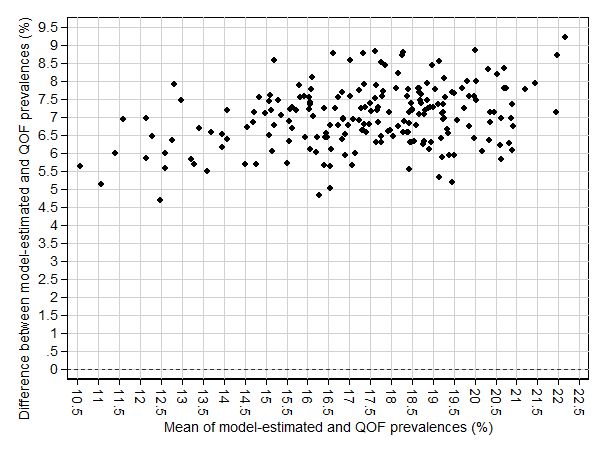
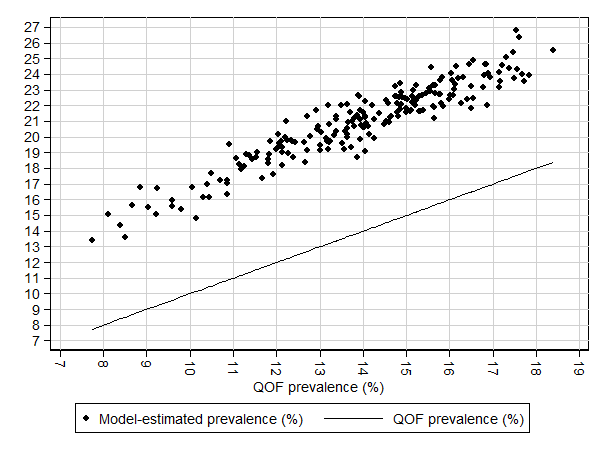


Figure 5 is a scatter plot of CCG-level model-estimated and QOF prevalence of diagnosed hypertension. The disagreement analysis between model-estimated and QOF prevalences (%) of diagnosed hypertension in practices has a Kendall's tau-a of 0.628 (95% CI 0.619-0.638, p 2.2x10-308). The best way to display the data is to plot the difference between the measurements by the two methods for each subject against their mean, creating a Bland-Altman plot. The plot for the CCG-level local diagnosed hypertension prevalence (Figure 5) shows explicitly the lack of agreement. The difference between the estimates is close to 7%. This is plausible if (a) some diagnosed hypertension remits e.g. with lifestyle changes, so patients are taken off registers and/or (b) some patients with hypertension have a diagnosis made, but then change GP, or the current GP has told the patient they have hypertension but does not record it on the hypertension register (they may even be treated with antihypertensives).

Figure 5: scatter plot of CCG-level model-estimated and QOF prevalence of diagnosed hypertension



Kendall's tau-a between mean prevalence and prevalence difference of diagnosed hypertension at practice level is -0.056 (95%, CI (-0.072- -0.041, p=4.0x10-12). Figure 6 shows the Bland-Altman plot for the practice-level QOF and estimated prevalences for diagnosed hypertension with much wider variation. As for the CCG-level comparison the difference between the estimates is close to 7%. Table 23 shows the percentile differences between practice-level model-estimated and QOF prevalences of diagnosed hypertension.

Table 23: percentile differences between practice-level model-estimated and QOF prevalences of diagnosed hypertension

|  |  |  |  |
| --- | --- | --- | --- |
| *Percent* | *Percentile* | *(95%* | *CI)* |
| 0 | -7.1 | -7.1) | |
| 25 | 5.7 | (5.6 - 5.7) | |
| 50 | 6.9 | (6.9 - 7.0) | |
| 75 | 8.2 | (8.1 - 8.2) | |
| 100 | 16.0 | (16.0 - 16.0) | |

Figure 6: Bland-Altman plot for practice-level model-estimated and QOF prevalence of diagnosed hypertension

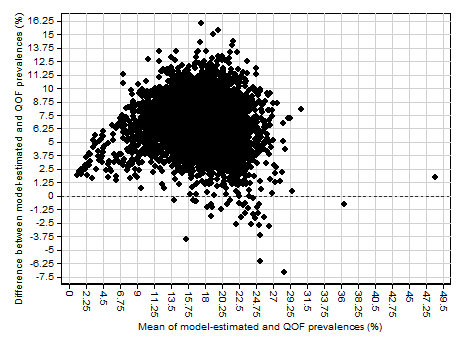


Figure 7 is a scatter plot of practice-level model-estimated and QOF prevalence of diagnosed hypertension. The Kendall tau-a between the mean of the two prevalences and the difference between the model-estimated and QOF prevalences was 0.184 (95% CI, 0.088 to 0.277; *P*=.00023), showing that prevalence means and model-QOF differences are 18.4 percent more likely to be concordant than to be discordant. This indicates that model-estimated prevalences have a higher variability than QOF prevalences, as suggested by the slight upwards trend in the Bland-Altman plot. Table 22 shows percentile differences between model-estimated and QOF prevalence of diagnosed hypertension. Note that all percentile differences are positive, even percentile zero, implying that every model-based prevalence was greater than the corresponding QOF prevalence for the same CCG.

Figure 7: scatter plot of practice -level model-estimated and QOF prevalence of diagnosed hypertension

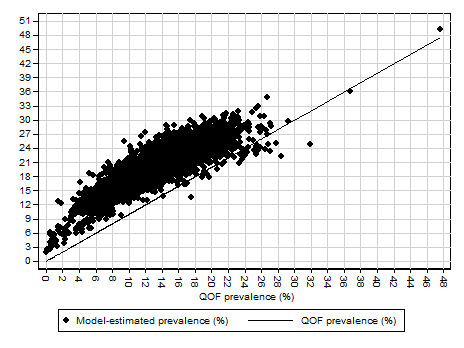


Table 24: comparison of aggregate local estimates with England Regions in derivation dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | North East | North West | Yorkshire & Humber | East Midlands | West Midlands | East England | London | South East Coast | South Central | South West | Total |
| Derivation dataset | **Population** |  |  |  |  |  |  |  |  |  |  |  |
| **Cases** |  |  |  |  |  |  |  |  |  |  |  |
| **Prevalence** |  |  |  |  |  |  |  |  |  |  |  |
| Local estimates | **Population** |  |  |  |  |  |  |  |  |  |  |  |
| **Cases** |  |  |  |  |  |  |  |  |  |  |  |
| **Prevalence** |  |  |  |  |  |  |  |  |  |  |  |
| Difference | |  |  |  |  |  |  |  |  |  |  |  |

# Discussion

## Summary

In order to generate local estimates for the prevalence of hypertension in England several models were fitted using data from the Health Survey for England collected in 2013 and 2014. After conducting a literature review to inform the analysis, all the known risk or protective factors available in HSfE were tested and either included or excluded from our models using Wald tests and likelihood ratios. Basing our decision in the performance of each model and the variables available at the local level, a “complete model” (the model with the best possible performance) and a “local model” (the model with the best performance only using variables locally available) were fitted for diagnosed hypertension and undiagnosed hypertension respectively. The information from the two local models and local data retrieved from different sources were used with the inverse probability weights method in order to obtain local prevalence estimates.

## Strengths and limitations

### Strengths

In terms of the national data source, strengths of this analysis include the use of a nationally-representative health survey, and rigorous measurement methods by trained survey nurses. HSfE also includes a broad range of risk factor data.

### Limitations

Several possible limitations should be acknowledged when considering the performance of this model, these are mostly in relation to the source of the data used to create the model (HSfE) and to generate the local estimates (Census data from 2011 and others).

In the first place, as with most household surveys, nonresponse may introduce volunteer bias. Furthermore, individuals living in nursing homes, prisons or military camps, who can display clearly different demographics, were not included in this survey. This is likely to have at the local level for some areas. In addition, the conditions in which the blood pressure measurements were performed might affect results obtained through variations such as the hour of the day in which the measurements took place, or situations such as respondents suffering from white-coat hypertension. Moreover, it is possible that other significant variables have been excluded from the complete models, due to the characteristics of the HSfE data. Variables for physical activity contained a high percentage of missing data, while fruit and vegetable consumption results for 2014 were significantly concentrated within one category.

As observed in our model comparison, the performance of the local models is inferior to the complete models. This is a result of some of the limitations of using local data, as not all the variables in the complete model where available locally in a valid format, or changes (e.g. in the differences between categories) were required in order to adapt our model to the available data. More importantly perhaps, interactions, which were observed to improve the performance of the complete models to some extent, could not be included in the final local models.

Another limitation arises from the use of local data, retrieved from public sources (like the 2011 Census) that do not provide annual updates. In addition to evident changes in population demographics happening over time, each year new postcodes are added, and the relation between LSOAs, MSOAs and GP practices potentially changes. This changes produce variations a mismatches when trying to link LSOA/MSOA information to practices.

# References

1. Poulter NR, Prabhakaran D, Caulfield M. Hypertension. The Lancet 2015;386(9995):801-12. Link: <http://dx.doi.org/10.1016/S0140-6736(14)61468-9>.

2. National Institute for Health and Care Excellence N. Hypertension in adults: diagnosis and management, 2011:38. <https://www.nice.org.uk/guidance/cg127>.

3. Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, White IR, Caulfield MJ, Deanfield JE, Smeeth L, Williams B, Hingorani A, Hemingway H. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. The Lancet 2014;383(9932):1899-911. Link: <http://www.sciencedirect.com/science/article/pii/S0140673614606851>.

4. Wang Y, Wang QJ. The Prevalence of Prehypertension and Hypertension Among US Adults According to the New Joint National Committee Guidelines: New Challenges of the Old Problem. Archives of Internal Medicine 2004;164(19):2126-34. Link: <http://archinte.ama-assn.org/cgi/content/abstract/164/19/2126>.

5. Wald NJ, Law MR. A strategy to reduce cardiovascular disease by more than 80%. BMJ 2003;326(7404):1419. Link: <http://www.bmj.com/cgi/content/abstract/326/7404/1419>.

6. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. The Lancet 2005;365(9455):217-23. Link: <http://www.sciencedirect.com/science/article/B6T1B-4F7Y352-11/2/25cd4894b7c8bce2136ca95ac5007fc6>.

7. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 2012;380(9859):2224-60. Link: <http://www.sciencedirect.com/science/article/pii/S0140673612617668>.

8. McAlister FA, Wilkins K, Joffres M, Leenen FHH, Fodor G, Gee M, Tremblay MS, Walker R, Johansen H, Campbell N. Changes in the rates of awareness, treatment and control of hypertension in Canada over the past two decades. Canadian Medical Association Journal 2011;183(9):1007-13. Link: <http://www.cmaj.ca/content/183/9/1007.abstract>.

9. Falaschetti E, Mindell J, Knott C, Poulter N. Hypertension management in England: a serial cross-sectional study from 1994 to 2011. The Lancet 2014;383(9932):1912-19. Link: <http://www.sciencedirect.com/science/article/pii/S0140673614606887>.

10. Peters SAE, Huxley RR, Woodward M. Comparison of the Sex-Specific Associations Between Systolic Blood Pressure and the Risk of Cardiovascular Disease: A Systematic Review and Meta-analysis of 124 Cohort Studies, Including 1.2 Million Individuals. Stroke 2013;44(9):2394-401. Link: <http://stroke.ahajournals.org/content/44/9/2394.abstract>.

11. Blacher J, Levy BI, Mourad J-J, Safar ME, Bakris G. From epidemiological transition to modern cardiovascular epidemiology: hypertension in the 21st century. The Lancet 2016. Link: <http://www.sciencedirect.com/science/article/pii/S0140673616000027>.

12. Lawes CMMa, Hoorn SVa, Law MRb, Elliott Pc, MacMahon Sd, Rodgers Aa. Blood pressure and the global burden of disease 2000. Part II: Estimates of attributable burden. Journal of Hypertension 2006;24(3):423-30. Link: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=ovfth&AN=00004872-200603000-00002>.

13. Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. The Lancet 2002;360(9349):1903-13. Link: <http://www.sciencedirect.com/science/article/B6T1B-47F6W3G-7/2/0236d6ada7642c69524eef30478e59c3>.

14. He FJ, MacGregor GA. A comprehensive review on salt and health and current experience of worldwide salt reduction programmes. J Hum Hypertens 2008;23(6):363-84. Link: <http://dx.doi.org/10.1038/jhh.2008.144>.

15. Whelton PK, He J, Muntner P. Prevalence, awareness, treatment and control of hypertension in North America, North Africa and Asia. J Hum Hypertens 2004;18(8):545-51. Link: <http://dx.doi.org/10.1038/sj.jhh.1001701>.

16. Hajat C, Tilling K, Stewart JA, Lemic-Stojcevic N, Wolfe CDA. Ethnic Differences in Risk Factors for Ischemic Stroke: A European Case-Control Study 5. Stroke 2004;35(7):1562-67. Link: <http://stroke.ahajournals.org/cgi/content/abstract/35/7/1562>.

17. Wolf-Maier K, Cooper RS, Banegas JR, Giampaoli S, Hense HW, Joffres M, Kastarinen M, Poulter N, Primatesta P, Rodriguez-Artalejo F, Stegmayr B, Thamm M, Tuomilehto J, Vanuzzo D, Vescio F. Hypertension Prevalence and Blood Pressure Levels in 6 European Countries, Canada, and the United States. JAMA: The Journal of the American Medical Association 2003;289(18):2363-69. Link: <http://jama.ama-assn.org/cgi/content/abstract/289/18/2363>.

18. Wang YR, Alexander GC, Stafford RS. Outpatient Hypertension Treatment, Treatment Intensification, and Control in Western Europe and the United States. Archives of Internal Medicine 2007;167(2):141-47. Link: <http://archinte.ama-assn.org/cgi/content/abstract/167/2/141>.

19. Moradi-Lakeh M, El Bcheraoui C, Tuffaha M, Daoud F, Al Saeedi M, Basulaiman M, Memish ZA, Al Mazroa MA, Al Rabeeah AA, Mokdad AH. The health of Saudi youths: current challenges and future opportunities. BMC Fam Pract 2016;17(1):1-9. Link: <http://dx.doi.org/10.1186/s12875-016-0425-z>.

20. Wu J, Cheng X, Qiu L, Xu T, Zhu G, Han J, Xia L, Qin X, Cheng Q, Liu Q. Prevalence and Clustering of Major Cardiovascular Risk Factors in China: A Recent Cross-Sectional Survey. Medicine (Baltimore) 2016;95(10):e2712. Link: <http://journals.lww.com/md-journal/Fulltext/2016/03080/Prevalence_and_Clustering_of_Major_Cardiovascular.8.aspx>.

21. Yoon SS, Carroll MD, Fryar CD. Hypertension Prevalence and Control Among Adults: United States, 2011-2014. NCHS data brief 2015(220):1-8. Link: <http://www.cdc.gov/nchs/products/databriefs/db220.htm>

22. Joffres MR, Ghadirian P, Fodor JG, Petrasovits A, Chockalingam A, Hamet P. Awareness, Treatment, and Control of Hypertension in Canada. American Journal of Hypertension 1997;10(10):1097-102. Link: <http://www.sciencedirect.com/science/article/B6T0Y-3Y9GDPJ-148/2/09f6b5e0d4314dfc55665a5d316a71e1>.

23. Joffres M, Falaschetti E, Gillespie C, Robitaille C, Loustalot F, Poulter N, McAlister FA, Johansen H, Baclic O, Campbell N. Hypertension prevalence, awareness, treatment and control in national surveys from England, the USA and Canada, and correlation with stroke and ischaemic heart disease mortality: a cross-sectional study. BMJ Open 2013;3(8). Link: <http://bmjopen.bmj.com/content/3/8/e003423.abstract>.

24. Brindel P, Hanon O, Dartigues J-F, Ritchie K, Lacombe J-M, Ducimetière P, Alpérovitch A, Tzourio C, Investigators ftCS. Prevalence, awareness, treatment, and control of hypertension in the elderly: the Three City study. Journal of Hypertension 2006;24(1):51-58. Link: <http://journals.lww.com/jhypertension/Fulltext/2006/01000/Prevalence,_awareness,_treatment,_and_control_of.11.aspx>.

25. MacDonald TM, Morant SV. Prevalence and treatment of isolated and concurrent hypertension and hypercholesterolaemia in the United Kingdom. British Journal of Clinical Pharmacology 2008;65(5):775-86. Link: <http://dx.doi.org/10.1111/j.1365-2125.2007.03072.x>.

26. Leenen FHH, Dumais J, McInnis NH, Turton P, Stratychuk L, Nemeth K, Moy Lum-Kwong M, Fodor G. Results of the Ontario Survey on the Prevalence and Control of Hypertension. Canadian Medical Association Journal 2008;178(11):1441-49. Link: <http://www.cmaj.ca/content/178/11/1441.abstract>.

27. Neuhauser HK, Adler C, Rosario AS, Diederichs C, Ellert U. Hypertension prevalence, awareness, treatment and control in Germany 1998 and 2008-11. J Hum Hypertens 2015;29(4):247-53. Link: <http://dx.doi.org/10.1038/jhh.2014.82>.

28. Thorpe Jr RJ, Brandon DT, LaVeist TA. Social context as an explanation for race disparities in hypertension: Findings from the Exploring Health Disparities in Integrated Communities (EHDIC) Study. Social Science & Medicine 2008;67(10):1604-11. Link: <http://www.sciencedirect.com/science/article/pii/S0277953608003560>.

29. Agyemang C, Bhopal R. Is the blood pressure of people from African origin adults in the UK higher or lower than that in European origin white people? A review of cross-sectional data. J Hum Hypertens 2003;17(8):523-34. Link: <http://dx.doi.org/10.1038/sj.jhh.1001586>.

30. Agyemang C, Bhopal R, Redekop WK. Does the pulse pressure in people of European, African and South Asian descent differ? A systematic review and meta-analysis of UK data. J Hum Hypertens 2007;21(8):598-609. Link: <http://dx.doi.org/10.1038/sj.jhh.1002191>.

31. Primatesta P, Bost L, Poulter NR. Blood pressure levels and hypertension status among ethnic groups in England. J Hum Hypertens 2000;14(2):143-8. Link.

32. Modesti PA, Reboldi G, Cappuccio FP, Agyemang C, Remuzzi G, Rapi S, Perruolo E, Parati G. Panethnic Differences in Blood Pressure in Europe: A Systematic Review and Meta-Analysis. PloS one 2016;11(1):e0147601. Link.

33. Lyratzopoulos G, McElduff P, Heller RF, Hanily M, Lewis PS. Comparative levels and time trends in blood pressure, total cholesterol, Body Mass Index and smoking among Caucasian and South-Asian participants of a UK primary-care based cardiovascular risk factor screening programme. BMC Public Health 2005;5(1):1-9. Link: <http://dx.doi.org/10.1186/1471-2458-5-125>.

34. Droyvold WB, Midthjell K, Nilsen TIL, Holmen J. Change in body mass index and its impact on blood pressure: a prospective population study. Int J Obes Relat Metab Disord 2005;29(6):650-55. Link: <http://dx.doi.org/10.1038/sj.ijo.0802944>.

35. Colhoun HM, Hemingway H, Poulter NR. Socio-economic status and blood pressure: an overview analysis. J Hum Hypertens 1998;12(2):91-110. Link: <http://www.nature.com/jhh/journal/v12/n2/pdf/1000558a.pdf?origin=publication_detail>.

36. Brummett BH, Babyak MA, Siegler IC, Shanahan M, Harris KM, Elder GH, Williams RB. Systolic Blood Pressure, Socioeconomic Status, and Biobehavioral Risk Factors in a Nationally Representative US Young Adult Sample. Hypertension 2011;58(2):161-66. Link: <http://hyper.ahajournals.org/content/58/2/161.abstract>.

37. Conen D, Glynn RJ, Ridker PM, Buring JE, Albert MA. Socioeconomic status, blood pressure progression, and incident hypertension in a prospective cohort of female health professionals. Eur Heart J 2009;30(11):1378-84. Link.

38. Brennan-Olsen SL, Williams LJ, Holloway KL, Hosking SM, Stuart AL, Dobbins AG, Pasco JA. Small area-level socioeconomic status and all-cause mortality within 10 years in a population-based cohort of women: Data from the Geelong Osteoporosis Study. Preventive medicine reports 2015;2:505-11. Link: <http://hyper.ahajournals.org/content/39/3/772.long>.

39. Matthews KA, Kiefe CI, Lewis CE, Liu K, Sidney S, Yunis C. Socioeconomic trajectories and incident hypertension in a biracial cohort of young adults. Hypertension 2002;39(3):772-6. Link: <http://hyper.ahajournals.org/content/39/3/772.long>.

40. Agyemang C, Addo J, Bhopal R, de Graft Aikins A, Stronks K. Cardiovascular disease, diabetes and established risk factors among populations of sub-Saharan African descent in Europe: a literature review. Globalization and Health 2009;5(1):1-17. Link: <http://dx.doi.org/10.1186/1744-8603-5-7>.

41. Diez Roux AV, Chambless L, Merkin SS, Arnett D, Eigenbrodt M, Nieto FJ, Szklo M, Sorlie P. Socioeconomic Disadvantage and Change in Blood Pressure Associated With Aging. Circulation 2002;106(6):703-10. Link: <http://circ.ahajournals.org/cgi/content/abstract/106/6/703>.

42. Landsbergis PA, Diez-Roux AV, Fujishiro K, Baron S, Kaufman JD, Meyer JD, Koutsouras G, Shimbo D, Shrager S, Stukovsky KH, Szklo M. Job Strain, Occupational Category, Systolic Blood Pressure, and Hypertension Prevalence: The Multi-Ethnic Study of Atherosclerosis. Journal of Occupational and Environmental Medicine 2015;57(11):1178-84. Link: <http://journals.lww.com/joem/Fulltext/2015/11000/Job_Strain,_Occupational_Category,_Systolic_Blood.6.aspx>.

43. Stamler J. The INTERSALT Study: background, methods, findings, and implications. The American journal of clinical nutrition 1997;65(2 Suppl):626s-42s. Link: <http://ajcn.nutrition.org/content/65/2/626S.long>.

44. Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP, Meerpohl JJ. Effect of lower sodium intake on health: systematic review and meta-analyses. Bmj 2013;346:f1326. Link: <http://www.bmj.com/content/346/bmj.f1326.long>.

45. He FJ, Pombo-Rodrigues S, Macgregor GA. Salt reduction in England from 2003 to 2011: its relationship to blood pressure, stroke and ischaemic heart disease mortality. BMJ Open 2014;4(4):e004549. Link: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3987732/>.

46. Aburto NJ, Hanson S, Gutierrez H, Hooper L, Elliott P, Cappuccio FP. Effect of increased potassium intake on cardiovascular risk factors and disease: systematic review and meta-analyses. Bmj 2013;346:f1378. Link: <http://www.bmj.com/content/346/bmj.f1378.long>.

47. Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM, Lin PH, Karanja N. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. N Engl J Med 1997;336(16):1117-24. Link: <http://www.nejm.org/doi/full/10.1056/NEJM199704173361601>.

48. He FJ, MacGregor GA. Fortnightly review: Beneficial effects of potassium. Bmj 2001;323(7311):497-501. Link: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1121081/>.

49. Cornelissen VA, Smart NA. Exercise Training for Blood Pressure: A Systematic Review and Meta‐analysis. Journal of the American Heart Association 2013;2(1). Link: <http://jaha.ahajournals.org/content/2/1/e004473.abstract>.

50. Hollingworth M, Harper A, Hamer M. Dose-response associations between cycling activity and risk of hypertension in regular cyclists: The UK Cycling for Health Study. J Hum Hypertens 2015;29(4):219-23. Link: <http://www.nature.com/jhh/journal/v29/n4/full/jhh201489a.html>.

51. Pavey TG, Peeters G, Bauman AE, Brown WJ. Does Vigorous Physical Activity Provide Additional Benefits beyond Those of Moderate? Medicine & Science in Sports & Exercise 2013;45(10):1948-55. Link: <http://journals.lww.com/acsm-msse/Fulltext/2013/10000/Does_Vigorous_Physical_Activity_Provide_Additional.13.aspx>.

52. Marchi KC, Muniz JJ, Tirapelli CR. Hypertension and chronic ethanol consumption: What do we know after a century of study? World journal of cardiology 2014;6(5):283-94. Link: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4062120/>.

53. Taylor B, Irving HM, Baliunas D, Roerecke M, Patra J, Mohapatra S, Rehm J. Alcohol and hypertension: gender differences in dose-response relationships determined through systematic review and meta-analysis. Addiction (Abingdon, England) 2009;104(12):1981-90. Link: <http://onlinelibrary.wiley.com/doi/10.1111/j.1360-0443.2009.02694.x/abstract;jsessionid=FA37CE1628F128F699D77E28D55CC070.f02t02>.

54. Briasoulis A, Agarwal V, Messerli FH. Alcohol consumption and the risk of hypertension in men and women: a systematic review and meta-analysis. Journal of clinical hypertension (Greenwich, Conn.) 2012;14(11):792-8. Link: <http://onlinelibrary.wiley.com/doi/10.1111/jch.12008/abstract>.

55. Maheswaran R, Gill JS, Davies P, Beevers DG. High blood pressure due to alcohol. A rapidly reversible effect. Hypertension 1991;17(6 Pt 1):787-92. Link: <http://hyper.ahajournals.org/content/17/6_Pt_1/787.abstract>.

56. Liang R, Zhang B, Zhao X, Ruan Y, Lian H, Fan Z. Effect of exposure to PM2.5 on blood pressure: a systematic review and meta-analysis. Journal of Hypertension 2014;32(11):2130-41. Link: <http://journals.lww.com/jhypertension/Fulltext/2014/11000/Effect_of_exposure_to_PM2_5_on_blood_pressure___a.4.aspx>.

57. Meng L, Chen D, Yang Y, Zheng Y, Hui R. Depression increases the risk of hypertension incidence: a meta-analysis of prospective cohort studies. J Hypertens 2012;30(5):842-51. Link.

58. Nabi H, Chastang JF, Lefevre T, Dugravot A, Melchior M, Marmot MG, Shipley MJ, Kivimaki M, Singh-Manoux A. Trajectories of depressive episodes and hypertension over 24 years: the Whitehall II prospective cohort study. Hypertension 2011;57(4):710-6. Link: <http://hyper.ahajournals.org/content/57/4/710.long>.

59. Shah MT, Zonderman AB, Waldstein SR. Sex and age differences in the relation of depressive symptoms with blood pressure. Am J Hypertens 2013;26(12):1413-20. Link: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3816627/>.

60. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. 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 2002;360:1903-13. Link.

61. Chen X, Wang Y. Tracking of Blood Pressure From Childhood to Adulthood. A Systematic Review and Meta-Regression Analysis. Circulation 2008:CIRCULATIONAHA. Link: <http://circ.ahajournals.org/cgi/content/abstract/CIRCULATIONAHA.107.730366v1>.

62. Hardy R, Kuh D, Whincup PH, Wadsworth ME. Age at puberty and adult blood pressure and body size in a British birth cohort study. Journal of Hypertension 2006;24(1):59-66. Link: <http://journals.lww.com/jhypertension/Fulltext/2006/01000/Age_at_puberty_and_adult_blood_pressure_and_body.12.aspx>.

63. Lane D, Beevers DG, Lip GYH. Ethnic differences in blood pressure and the prevalence of hypertension in England. Journal of Human Hypertension 2002;16(4):6. Link: <http://www.nature.com/jhh/journal/v16/n4/full/1001371a.html>.

64. Forman JP, Stampfer MJ, Curhan GC. DIet and lifestyle risk factors associated with incident hypertension in women. JAMA 2009;302(4):401-11. Link: <http://dx.doi.org/10.1001/jama.2009.1060>.

65. Doll S, Paccaud F, Bovet P, Burnier M, Wietlisbach V. Body mass index, abdominal adiposity and blood pressure: consistency of their association across developing and developed countries. Int J Obes Relat Metab Disord 2002;26(1):48-57. Link: <http://www.nature.com/ijo/journal/v26/n1/full/0801854a.html>.

66. Arabshahi S, Busingye D, Subasinghe AK, Evans RG, Riddell MA, Thrift AG. Adiposity has a greater impact on hypertension in lean than not-lean populations: a systematic review and meta-analysis. European journal of epidemiology 2014;29(5):311-24. Link: <http://dx.doi.org/10.1007/s10654-014-9911-6>.

67. He FJ, Li J, Macgregor GA. Effect of longer term modest salt reduction on blood pressure: Cochrane systematic review and meta-analysis of randomised trials. Bmj 2013;346:f1325. Link: <http://www.bmj.com/content/346/bmj.f1325.long>.

68. Kotchen TA, Cowley AW, Frohlich ED. Salt in Health and Disease — A Delicate Balance. New England Journal of Medicine 2013;368(13):1229-37. Link: <http://www.nejm.org/doi/full/10.1056/NEJMra1212606>.

69. Mente A, O'Donnell MJ, Rangarajan S, McQueen MJ, Poirier P, Wielgosz A, Morrison H, Li W, Wang X, Di C, Mony P, Devanath A, Rosengren A, Oguz A, Zatonska K, Yusufali AH, Lopez-Jaramillo P, Avezum A, Ismail N, Lanas F, Puoane T, Diaz R, Kelishadi R, Iqbal R, Yusuf R, Chifamba J, Khatib R, Teo K, Yusuf S. Association of urinary sodium and potassium excretion with blood pressure. N Engl J Med 2014;371(7):601-11. Link: <http://www.escardio.org/Congresses-&-Events/Congress-resources/ESC-Congress-365/PURE-Sodium-Heterogeneity-in-the-associations-of-urinary-sodium-and-potassium-w>.

70. Soardo G, Donnini D, Varutti R, Milocco C, Basan L, Esposito W, Casaccio D, Isola M, Soldano F, Sechi LA. Effects of alcohol withdrawal on blood pressure in hypertensive heavy drinkers. Journal of Hypertension 2006;24(8):1493-98. Link: <http://journals.lww.com/jhypertension/Fulltext/2006/08000/Effects_of_alcohol_withdrawal_on_blood_pressure_in.7.aspx>.

71. Steffens AA, Moreira LB, Fuchs SC, Wiehe M, Gus M, Fuchs FD. Incidence of hypertension by alcohol consumption: is it modified by race? Journal of Hypertension 2006;24(8):1489-92. Link: <http://journals.lww.com/jhypertension/Fulltext/2006/08000/Incidence_of_hypertension_by_alcohol_consumption_.6.aspx>.

72. D'Elia L, De Palma D, Rossi G, Strazzullo V, Russo O, Iacone R, Fazio V, Strazzullo P, Galletti F. Not smoking is associated with lower risk of hypertension: results of the Olivetti Heart Study. European journal of public health 2014;24(2):226-30. Link: <http://eurpub.oxfordjournals.org/content/24/2/226.long>.

73. Halimi J-M, Giraudeau B, Vol S, Cacès E, Nivet H, Tichet J. The risk of hypertension in men: direct and indirect effects of chronic smoking. Journal of Hypertension 2002;20(2):187-93. Link: <http://journals.lww.com/jhypertension/Fulltext/2002/02000/The_risk_of_hypertension_in_men__direct_and.7.aspx>.

74. John U, Meyer C, Hanke M, Volzke H, Schumann A. Smoking status, obesity and hypertension in a general population sample: a cross-sectional study. QJM : monthly journal of the Association of Physicians 2006;99(6):407-15. Link: <http://qjmed.oxfordjournals.org/content/99/6/407.long>.

75. Pankova A, Kralikova E, Fraser K, Lajka J, Svacina S, Matoulek M. No difference in hypertension prevalence in smokers, former smokers and non-smokers after adjusting for body mass index and age: a cross-sectional study from the Czech Republic, 2010. Tobacco induced diseases 2015;13(1):24. Link: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4531506/>.

76. Linneberg A, Jacobsen RK, Skaaby T, Taylor AE, Fluharty ME, Jeppesen JL, Bjorngaard JH, Asvold BO, Gabrielsen ME, Campbell A, Marioni RE, Kumari M, Marques-Vidal P, Kaakinen M, Cavadino A, Postmus I, Ahluwalia TS, Wannamethee SG, Lahti J, Raikkonen K, Palotie A, Wong A, Dalgard C, Ford I, Ben-Shlomo Y, Christiansen L, Kyvik KO, Kuh D, Eriksson JG, Whincup PH, Mbarek H, de Geus EJ, Vink JM, Boomsma DI, Smith GD, Lawlor DA, Kisialiou A, McConnachie A, Padmanabhan S, Jukema JW, Power C, Hypponen E, Preisig M, Waeber G, Vollenweider P, Korhonen T, Laatikainen T, Salomaa V, Kaprio J, Kivimaki M, Smith BH, Hayward C, Sorensen TI, Thuesen BH, Sattar N, Morris RW, Romundstad PR, Munafo MR, Jarvelin MR, Husemoen LL. Effect of Smoking on Blood Pressure and Resting Heart Rate: A Mendelian Randomization Meta-Analysis in the CARTA Consortium. Circulation. Cardiovascular genetics 2015;8(6):832-41. Link: <http://circgenetics.ahajournals.org/content/8/6/832.long>.

77. Fan AZ, Strasser SM, Zhang X, Fang J, Crawford CG. State socioeconomic indicators and self-reported hypertension among US adults, 2011 behavioral risk factor surveillance system. Preventing chronic disease 2015;12:E27. Link: <http://www.cdc.gov/pcd/issues/2015/14_0353.htm>.

78. Ezzati M, Oza S, Danaei G, Murray CJL. Trends and Cardiovascular Mortality Effects of State-Level Blood Pressure and Uncontrolled Hypertension in the United States. Circulation 2008;117(7):905-14. Link: <http://circ.ahajournals.org/cgi/content/abstract/117/7/905>.

79. Health FoP. FPHM's hypertension prevalence model. Link: <http://www.fphm.org.uk/resources/AtoZ/toolkit_hypertension/tools.asp>.

80. Observatories AoPH. Disease prevalence models: CHD, hypertension, stroke and COPD. 2008: Association of Public Health Observatories. Link: <http://www.apho.org.uk/resource/view.aspx?RID=39389>.

81. Gary Boodhna SB, Robin Darton, Sue Faulding,Catherine Henderson, Jennifer Mindell, Alison Moody, Linda Ng Fat,Caireen Roberts, Alice Ryley, Shaun Scholes, Joanne Thompson,Laura Weston, Raphael Wittenberg. Health Survey for England 2013: Methods and documentation. 2014:210. Link: [www.hscic.gov.uk/pubs/hse2013](http://www.hscic.gov.uk/pubs/hse2013).

82. Sally Bridges RD, Sara Evans-Lacko, Elizabeth Fuller,, Claire Henderson NI, Natalie Maplethorpe, Jennifer Mindell,, Alison Moody LNF, Keeva Rooney, Nina Sal, Rachel Scantlebury,, Shaun Scholes GT, Raphael Wittenberg. Health Survey for England 2014: Methods and documentation. 2015. Link: <http://www.hscic.gov.uk/catalogue/PUB19295/HSE2014-Methods-and-docs.pdf>.

83. Li P, Stuart EA, Allison DB. Multiple imputation: A flexible tool for handling missing data. JAMA 2015;314(18):1966-67. Link: <http://dx.doi.org/10.1001/jama.2015.15281>.

84. He Y, Zaslavsky AM, Landrum MB, Harrington DP, Catalano P. Multiple imputation in a large-scale complex survey: a practical guide. Statistical Methods in Medical Research 2010;19(6):653-70. Link: <http://smm.sagepub.com/content/19/6/653.abstract>.

85. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338(jun29\_1):b2393. Link: <http://www.bmj.com>

<http://www.bmj.com/cgi/content/full/338/jun29_1/b2393>.

86. Kirkwood BR, Sterne JAC. Regression modelling. In: K M, editor. *Medical Statistics*. USA: Blackwell Publishing company, 2003:339-42.

87. Newson RB. Attributable and unattributable risks and fractions and other scenario comparisons. Stata Journal 2013;13(4):672-98. Link: <http://www.stata-journal.com/article.html?article=st0314>.

88. Newson RB. Scenario comparisons: How much good can we do?, 2014.

89. Ezzati M, Vander Hoorn S, Rodgers A, Lopez AD, Mathers CD, Murray CJL. Estimates of global and regional potential health gains from reducing muliple major risk factors. The Lancet 2003;362(9380):271-80. Link: <http://www.sciencedirect.com/science/article/pii/S0140673603139682>.

90. Neave A. Health Survey for England 2014: Trend Tables Commentary. 2015. Link: <http://www.hscic.gov.uk/catalogue/PUB19297/HSE2014-Trend-commentary.pdf>.

91. Mansournia MA, Altman DG. Inverse probability weighting. BMJ 2016;352. Link: <http://www.bmj.com/bmj/352/bmj.i189.full.pdf>.

92. Newson R. Confidence intervals for rank statistics: Somers' *D* and extensions. Stata J 2006;6(3):309-34. Link: <Go to ISI>://WOS:000240514400002.

93. Newson R. Confidence intervals for rank statistics: Percentile slopes, differences, and ratios. Stata J 2006;6(4):497-520. Link: <Go to ISI>://WOS:000242923600004.

# Annex 1: ROC curves

Figure 8: ROC curve for model DH0



Figure 9: ROC curve for model DH1



Figure 10: ROC curve for model DH2



Figure 11: ROC curve for model DH3



Figure 12: ROC curve for model DH 4



Figure 13: ROC curve for model DH 5



Figure 14: ROC curve for model DH 6



Figure 15: ROC curve for model DH 7



Figure 16: ROC curve for model UDH 0



Figure 17: ROC curve for model UDH 1





1. Figures in brackets are 95% confidence intervals [↑](#footnote-ref-1)