**Heart failure prevalence models for small populations:**

**Technical Document produced for Public Health England**

**Martin Cowie, Saba Mian, Jessica Morris, Roger Newson, Michael Soljak, Bowen Su**

**Department Primary Care & Public Health**

**School of Public Health**



**Contents**

[1 Executive Summary 4](#_Toc474152582)

[2 Background 5](#_Toc474152583)

[2.1 HF Risk Factors 5](#_Toc474152584)

[2.1.1 CHD and MI 6](#_Toc474152585)

[2.1.2 Hypertension 6](#_Toc474152586)

[2.1.3 Smoking 6](#_Toc474152587)

[2.1.4 Atrial Fibrillation 7](#_Toc474152588)

[2.1.5 Diabetes 7](#_Toc474152589)

[2.1.6 Age 7](#_Toc474152590)

[2.1.7 Congenital Heart Disease 7](#_Toc474152591)

[2.1.8 BMI 7](#_Toc474152592)

[2.1.9 Ethnicity 7](#_Toc474152593)

[2.1.10 Gender 8](#_Toc474152594)

[2.1.11 Alcohol Consumption 8](#_Toc474152595)

[2.1.12 Sleep Apnoea 8](#_Toc474152596)

[2.2 HF incidence and prevalence 13](#_Toc474152597)

[2.2.1 Quality & Outcomes Framework Data 14](#_Toc474152598)

[3 Methods 15](#_Toc474152599)

[3.1 HF prevalence from UK primary care and hospital data: Clinical Practice Research Datalink & Hospital Episode Statistics 15](#_Toc474152600)

[3.1.1 What is the Clinical Practice Research Datalink? 15](#_Toc474152601)

[3.1.2 Identification of HF cases and code lists for CPRD and HES doctor diagnosed cases 15](#_Toc474152602)

[3.1.3 Additional HF cases from clinical algorithm 17](#_Toc474152603)

[3.1.4 Additional HF cases from drugs algorithm 18](#_Toc474152604)

[3.1.5 CPRD risk factors 20](#_Toc474152605)

[3.1.6 CPRD HF Control group 20](#_Toc474152606)

[3.1.7 CPRD descriptive analyses 20](#_Toc474152607)

[3.1.8 CPRD prevalence and incidence 20](#_Toc474152608)

[3.1.9 1.1.1 CPRD regression modelling 21](#_Toc474152609)

[3.1.10 Interactions 21](#_Toc474152610)

[3.1.11 Internal validation 21](#_Toc474152611)

[3.1.12 Local prevalence estimates 21](#_Toc474152612)

[3.2 Validation of local estimates 25](#_Toc474152613)

[3.2.1 Internal validation 25](#_Toc474152614)

[3.2.2 External validation 25](#_Toc474152615)

[4 Results 28](#_Toc474152616)

[4.1 HF prevalence from CPRD 28](#_Toc474152617)

[4.1.1 Baseline characteristics of CPRD respondents 28](#_Toc474152618)

[4.1.2 Missing values 29](#_Toc474152619)

[4.1.3 CPRD prevalence and incidence 29](#_Toc474152620)

[4.1.4 Logistic regression modelling 33](#_Toc474152621)

[4.2 External validation – Bland Altman plots 34](#_Toc474152622)

[5 References 36](#_Toc474152623)

**HF prevalence model Technical Document**

# Executive Summary

Approximately 1 to 2% of the UK population is affected with HF, with prevalence rates expected to increase over the next decade.[1 ,2] Over 1 million individuals in the UK have HF and this places a significant financial burden on the NHS.[3 ,4] A recent survival analysis carried out using UK primary care records from The Health Improvement Network between 1998 and 2012 showed no improvement in survival over this period.[10] However early treatment can reduce significant mortality and morbidity associated with HF.[3]

We used linked Clinical Practice Research Datalink (CPRD) primary care electronic health records (EHRs) and Hospital Episode Statistics (HES) data as a robust data source for the identification of HF prevalence in patients who have been in contact with the health system. A comprehensive list of Medcodes, Prodcodes and ICD-10 codes was compiled for the identification of the CPRD and HES doctor diagnosed HF cases, and for the undiagnosed (but diagnosable) clinical and drugs algorithm HF cases. We used the literature review described in the Background to extract CPRD data on risk factors. We fitted uni-variate then multivariate logistic regression models for HF as described in previous publications, to produce odds ratios (ORs) and regression coefficients. Derived ORs (or regression coefficients) are used to estimate prevalence in small population subgroups. Having estimated the regression model parameters, we used these for out-of-sample prediction of HF prevalence. We also carried out internal and external validation.

We identified 224,265 HF cases from CPRD GP diagnoses. We found 14,097 additional HES diagnosed HF cases, 32,846 additional drugs algorithm probable HF cases, and no additional clinical algorithm probable HF cases. The prevalence of HF is estimated at around 0.74% for doctor-diagnosed HF, rising if algorithm cases are included as above. Between 2000-2015 incidence rates have halved and we have also shown a decrease in HF admission rates over this period. For the local estimates, Bland-Altman plots for the practice-level QOF and estimated prevalence for HF showed higher modelled prevalence compared with QOF registered prevalence. Scatter plots of practice-level model-estimated and QOF prevalence of diagnosed HF showed GP-diagnosed prevalence is lower than CPRD prevalence.

# Background

Heart failure (HF) is a chronic disease in which cardiac structure or function is abnormal, resulting in the inability of the heart to deliver oxygen at a rate that is necessary for metabolising tissues, regardless of normal filling pressures.[5 ,6] Approximately 1 to 2% of the UK population is affected with HF, with prevalence rates expected to increase over the next decade.[1 ,2] Over 1 million individuals in the UK have HF and this places a significant financial burden on the NHS.[3 ,4] In the USA, the estimated annual cost of HF in 2010 is estimated to be $39.2 billion or ∼2% of the total US health-care budget. Evaluations from different European countries indicate a similar share of HF-related costs in relation to overall health-care expenditure.[7] The overall global economic cost of HF in 2012 was estimated at $108 billion per annum.[8]

Treatment can reduce significant mortality and morbidity associated with HF.[3] However, almost half of HF patients die within five years of being diagnosed.[9] A recent survival analysis carried out using UK primary care records from The Health Improvement Network (THIN) between 1998 and 2012 showed no improvement in survival over this period.[10] Therefore, HF represents a major public health burden which is growing, but unlike coronary heart disease (CHD), this is not because of improved survival.[11]

Because HF is a medical diagnosis that patients may not report, using data from a single national survey such as the Health Survey for England would not provide a suitable basis for a HF prevalence model. However, linked Clinical Practice Research Datalink (CPRD) primary care electronic health records (EHRs) and Hospital Episode Statistics (HES) data should form a robust data source for the identification of HF patients who have been in contact with the health system. We therefore used these as primary data sources for prevalence modelling, with an adjustment if needed from published population surveys or from the National Heart Failure Audit,[12] data from which PCPH already holds.

## HF Risk Factors

Many studies have investigated the relationship between HF and risk factors such as history of CHD, hypertension, hypercholesterolemia, obesity, smoking, and arrhythmias.[13 ,14] [15-17] HFR risk prediction equations developed from UK primary care EHRs include age, body mass index (BMI), systolic blood pressure (BP), cholesterol/high density lipoprotein (HDL) ratio, HbA1c, material deprivation, ethnicity, smoking, duration and type of diabetes, atrial fibrillation (AF), cardiovascular disease (CVD), and chronic kidney disease (CKD).[18] Myocardial infarction (MI) has been found to be a precursor of HF and death, and is the most occurring event during post-MI patient management.[19] We conducted a non-systematic literature search to validate these predictors. HF risk factors we confirmed are shown in the following table, with associated references (**Table 1**).

Table : HF risk factor list

| Risk factor | References |
| --- | --- |
| MI | [19], [20], [21] & [22]  |
| CHD | [23], [19] [20], [21], [22] & [24]  |
| Congenital heart disease | [25] |
| Age | [26], [23], [27] & [28] |
| Hypertension | [11], [26], [29], [30], [31], [32] & [17] |
| Diabetes | [11], [33] & [34]  |
| Body Mass Index (BMI) | [11], [16] & [35]  |
| Smoking | [36], [13], [14], [11], [16], [37]  |
| Elevated low-density lipoprotein (LDL) cholesterol | [11] |
| Alcohol | [38] & [39]  |
| Sleep apnoea | [40] |
| Atrial fibrillation  | [16] |
| Ethnicity | [28] |
| Gender | [28] |

### CHD and MI

In developed countries, CHD is the leading cause of HF.[23] Chronic HF is the terminal stage of several cardiovascular diseases and cardiac dysfunction syndrome.[24] There has been an increase in the prevalence and incidence of HF in the past 25 years, which may be due to an improvement in the survival of patients after acute MI.[23] In the past 40 years in the United States, the odds of a previous MI being the cause of HF has increased by 26% in men and 48% in women per decade.[23] MI is the main clinical intermediate between CHD and HF [27]. The occurrence of HF post MI depends on the location and size of the infarct, the severity of artery disease and the development of ischemic mitral regurgitation.[27] In a population-based study, of 1,915 patients with a previous MI, 41% developed HF over a seven year follow-up period.[20]

### Hypertension

Studies have shown that the risk of HF is significantly reduced in hypertensive patients with a reduction in (systolic) BP. A wide range of anti-hypertensive drugs such as angiotensin-converting enzyme inhibitors (ACEi) and angiotensin receptor blockers (ARB) lower the risks of major cardiovascular events in hypertensive people.[41] Corrao et al (2015) investigated the relationship between long-term use of prescribed anti-hypertensive medication and the risk of hospitalisation (first occurrence) for HF, by carrying out a population-based case-control study [26]. Out of 76,017 patients, 622 patients were admitted to hospital for HF. These 622 patients were matched to 3,110 controls. The mean age of the patients and controls was 67 years and 54% of the patients were men. A 34% (95% CI 17%-48%) reduction in the risk of hospitalisation due to HF was observed with high adherence to antihypertensive medication.

Turbull et al (2007) found a significant 27% reduction in the risk of HF with ACEi treatment and a 12% reduction with ARB treatment for each 5mmHg reduction in blood pressure [31]. In an meta-analysis of 20 blood pressure controlling randomised controlled trials (RCTs), Zanchetti et al (2015) found a 46% (95% CI 36-55%) reduction in the risk of HF in hypertensive patients [30]. Chirinos et al (2015) found a significant relationship between late systolic hypertension and incident HF in the general population [32]. The adjusted hazard ratio for HF associated with hypertension was 1.27 (95% CI 1.18-1.36).

A study using data from the Physicians Health Study I (PHS I), a randomised, double-blind, placebo-controlled trial, found a significant relationship between systolic blood pressure (SBP) and the risk of HF [29]. Participants with untreated hypertensive SBP of 140-149 and >150mmHg, were at a 1.44 (95% CI 1.12-1.88, p <0.001) and 1.47 (95% CI 1.01-2.14, p <0.001), respectively, greater risk of HF in comparison to normotensive participants. Participants with treated hypertensive SBP of <140, 140-149 and ≥150mmHg, were at a 2.33 (95% CI 1.83-2.95, p <0.001), 1.98 (95% CI 1.51-2.60, p <0.001) and 2.77 (95% CI 2.06-3.73, p <0.001) respectively, greater risk of HF compared to normotensive participants.

### Smoking

Ahmed et al (2015) found a significant effect of smoking on the risk of HF [37]. Current smokers (n=629) were at a higher risk of HF compared to never smokers (n=2556) (adjusted HR 1.49, 95% CI 1.23 to 1.81, p<0.001). Former smokers (n=312) with a ≥32 pack year history were also associated with a significantly increased risk of HF compared to never smokers (adjusted HR 1.45, 95% CI 1.15 to 1.83, p=0.002). However, there was no significant relationship between former smokers with a ≤31 pack year history and HF risk. Gopal et al (2012) found that over a 9.4 year follow-up period of 2,125 participants from the Health, Aging, and Body Composition Study, 231 participants developed HF.[36] The study found that non-smokers had a HF incidence of 11.4 events per 1,000 person-years, former smokers had a HF incidence of 15.2 events per 1,000 person-years and current smokers had a HF incidence of 21.9 events per 1,000 person-years. Adjusted hazard ratios showed that current smokers (n=221) had a significantly greater risk of HF compared to never smokers (n=1165) (adjusted HR 1.73, 95% CI 1.15 to 2.59, p= 0.008). However, the adjusted hazard ratios did not show a strong association between former smoking (n=739) and risk of HF versus never smoking.

### Atrial Fibrillation

Atrial fibrillation and congestive HF often occur together and each condition can predispose to the other [42].

### Diabetes

Type 2 diabetes has been identified as a risk factor for HF.[43] HF is often a cause of disability and mortality in patients with type 2 diabetes mellitus, [34] with the risk of HF increasing considerably each year in diabetic patients [44]. Wang et al (2015) found a significant positive relationship between type 2 diabetes mellitus and incident HF [34]. Resengren et al (2015) found a strong positive association between type I diabetes and HF [33]. A four-fold increased risk of HF was observed in diabetic patients compared with population-based controls (HR 4.69, 95% CI 3.64 to 6.04, p< 0.001). Although most studies investigating HF risk in patients with diabetes have a sample of older patients with type 2 diabetes, a recent nationwide study of 33,402 people with type 1 diabetes in Sweden found a four-times greater risk of being admitted to hospital for HF compared to those without type 1 diabetes.[33] HbA1c was identified as a significant risk factor; for each 1% increase in HbA1c a 13% increase in the HF risk was observed.[33]

### Age

In the older population, HF is the most common cause of hospitalisation, hospital readmission and mortality.[26] Over a median follow-up of 11.2 years in the Multi-Ethnic Study of Atherosclerosis (MESA), 111 subjects out of 6,781 participants developed HF with preserved ejection fraction (HFpEF).[28] Age was found to be a significant risk factor for HFpEF, with increasing age (per SD) the risk increased by 2.27 (95% CI, 1.72-3.01, p<0.001).

### Congenital Heart Disease

Patients with congenital heart disease are increasingly reaching adulthood with the improvements in paediatric cardiology and cardiac surgery, improving the survival rates significantly.[45] Nevertheless, patients with congenital heart disease face complications in adulthood such as arrhythmias, endocarditis and HF.[46] A cohort study of 10,808 patients with congenital heart disease, with a follow-up time of 21 years, 274 (2.5%) patients were admitted for HF.[45] A study using the 2007 Nationwide Inpatient Sample found patients with adult congenital heart disease (ACHD) with ventricular septal defect had 1.54 times greater odds of developing HF (95% CI, 1.31-1.81) compared to those without ventricular septal defect.[25] The prevalence of congenital heart disease was <1% in the UK during 2000 to 2005, [47] therefore congenital heart disease will not make a large contribution to population prevalence of HF, and so it has not been included as a risk factor in the model.

### BMI

In a population-based cohort study of Danish male recruits (n = 12,850), 107 individuals were diagnosed with congestive HF. The 36-year risk of HF was 2.49 times greater in overweight individuals (95% CI, 1.56-3.99) and 4.52 times greater in obese individuals (95% CI 1.96-10.43) compared to those of normal weight (BMI 18.5 – 24.9kg/m2).[35]

### Ethnicity

In the MESA study, 111 subjects out of 6,781 participants developed HFpEF over a median follow-up of 11.2 years.[28] Black participants in the study were at a lower risk of HFpEF compared to White participants (95% CI, 0.26-0.82, p=0.009). However, there was no significant difference in the risk of HFpEF in Chinese and Hispanic participants compared to White participants. In their individual predictive risk model for HF using UK primary care EHRs, Hippisley-Cox et al included risk factors for age, BMI, systolic blood pressure, cholesterol/ high-density lipoprotein (HDL) ratio, glycosylated haemoglobin (HbA1c), material deprivation, ethnicity, smoking, diabetes duration, type of diabetes, atrial fibrillation, CVD, CKD, and family history of premature CHD.[18]

### Gender

In the MESA study, no significant difference was observed in the risk of developing HF between males and females. However, a study investigating the gender difference in the incidence and risk of new-onset heart HF in subjects from the Renal and Vascular Endstage Disease (PREVEND) study, found females were less likely to have new-onset of HF with reduce ejection fraction (HFrEF) (subhazard ratio = 0.47, 95 % CI 0.29–0.76, p = 0.002) compared to males. [48] However, women were more likely to have new-onset of HFpEF (subhazard ratio = 2.16, 95 % CI 1.21–3.83, p = 0.009) compared to males.

### Alcohol Consumption

In a prospective study of 3,530 males (60-79 years of age) without previous diagnosis of HF or MI, 198 HF cases were identified over the follow-up period of 11 years.[49] Male subjects drinking 35 drinks or more per week were at a 1.9 times greater risk of HF compared non-drinkers (95% CI, 1.02-3.54).

### Sleep Apnoea

In a population-based cohort study, Sleep and Health in Women, obstructive sleep apnoea symptoms such as snoring and excessive daytime sleepiness (EDS) were significantly associated with increased risk of HF. [40] Women who were snorers with EDS were at a higher risk of developing HF compared to non-snorers without EDS (HR 2.19, 95% CI, 1.14-4.18).

Table 2 summarises HF risk factors with their pooled, matched or adjusted odds ratios.

Table : HF risk factors with their pooled, matched or adjusted odds ratios

| Risk factor | Type of Odds Ratio | Odds Ratio | 95% CI | Effect on Outcome |
| --- | --- | --- | --- | --- |
| Hypertension |  |  |  |  |
| Antihypertensive therapy – High adherence  | Matched RR [26] | 0.66 | [0.52-0.83] | Reduced risk of HF (hypertension= risk factor) |
| ACEI therapy  | Crude RR [31] |  |  |  |
| ARB therapy  | Crude RR [31] |  |  |  |
| Intentional BP-lowering trials in hypertensive patients | Standardized RR [30] | 0.54 | [0.45-0.64] | Reduced risk of HF |
| Late systolic hypertension  | HR Adjusted for age, ethnicity, gender and heart rate [32] | 1.27 | [1.18-1.36] | Risk factor |
| Systolic blood pressure (mm/Hg) |  |  |  |  |
| Normotensive  |  |  |  |  |
| <120 | HR Adjusted for age, smoking, BMI, alcohol consumption, history of diabetes, history of atrial fibrillation, physical activity, egg intake, aspirin arm, and breakfast cereal intake [29] | 1.00 |  | Reference |
| 120-129 | HR Adjusted for age, smoking, BMI, alcohol consumption, history of diabetes, history of atrial fibrillation, physical activity, egg intake, aspirin arm, and breakfast cereal intake [29] | 1.10 | [0.89-1.37] | NS |
| 130-139 | HR Adjusted for age, smoking, BMI, alcohol consumption, history of diabetes, history of atrial fibrillation, physical activity, egg intake, aspirin arm, and breakfast cereal intake [29] | 1.35 | [1.09-1.68] | Risk factor |
| Hypertensive, Untreated |  |  |  |  |
| 140-149 | HR Adjusted for age, smoking, BMI, alcohol consumption, history of diabetes, history of atrial fibrillation, physical activity, egg intake, aspirin arm, and breakfast cereal intake [29] | 1.44 | [1.12-1.88] | Risk factor |
| >150 | HR Adjusted for age, smoking, BMI, alcohol consumption, history of diabetes, history of atrial fibrillation, physical activity, egg intake, aspirin arm, and breakfast cereal intake [29] | 1.47 | [1.01-2.14] | Risk factor |
| Hypertensive, Treated  |  |  |  |  |
| <140 | HR Adjusted for age, smoking, BMI, alcohol consumption, history of diabetes, history of atrial fibrillation, physical activity, egg intake, aspirin arm, and breakfast cereal intake [29] | 2.33 | [1.83-2.95] | Risk factor |
| 140-149 | HR Adjusted for age, smoking, BMI, alcohol consumption, history of diabetes, history of atrial fibrillation, physical activity, egg intake, aspirin arm, and breakfast cereal intake [29] | 1.98 | [1.51-2.60] | Risk factor |
| ≥150 | HR Adjusted for age, smoking, BMI, alcohol consumption, history of diabetes, history of atrial fibrillation, physical activity, egg intake, aspirin arm, and breakfast cereal intake [29] | 2.77 | [2.06-3.73] | Risk factor |
| MI |  |  |  |  |
| Infarct size  | Crude OR [19] | 1.067 | [1.036-1.099] | Risk factor |
| Age (per SD) | Crude OR [19]Multivariable Adjusted HR [28] | 1.052.27 | [1.005-1.097][1.72-3.01] | Risk factorRisk factor |
|  |  |  |  |  |
| Smoking |  |  |  |  |
| Never Smokers | HR Adjusted for age, gender, ethnicity [37] | 1.00 |  | Reference |
| Former Smokers, overall | HR Adjusted for age, gender, ethnicity [37] | 0.99  | [0.85-1.16] | NS |
| < 8 pack years | HR Adjusted for age, gender, ethnicity [37] | 0.91  | [0.71-1.17] | NS |
| 8-15 pack years | HR Adjusted for age, gender, ethnicity [37] | 0.71 | [0.51-0.98] |  |
| 16-31 pack years | HR Adjusted for age, gender, ethnicity [37] | 0.95 | [0.74-1.23] | NS |
| ≥32 pack years | HR Adjusted for age, gender, ethnicity [37] | 1.45 | [1.15-1.83] | Risk factor |
| Current Smokers | HR Adjusted for age, gender, ethnicity [37] | 1.49 | [1.23-1.81] | Risk factor |
| Never Smokers | HR Adjusted for age, coronary artery disease, left ventricular hypertrophy, systolic blood pressure, heart rate, albumin, fasting glucose, and creatinine [36] | 1.00 |  | Reference |
| Former Smokers | HR Adjusted for age, coronary artery disease, left ventricular hypertrophy, systolic blood pressure, heart rate, albumin, fasting glucose, and creatinine [36] | 1.31 | [0.98-1.75] | NS |
| Current smokers  | HR Adjusted for age, coronary artery disease, left ventricular hypertrophy, systolic blood pressure, heart rate, albumin, fasting glucose, and creatinine [36] | 1.73 | [1.15-2.59] | Risk factor |
| Diabetes |  |  |  |  |
| Type I | HR Adjusted for time-updated age, sex, time-updated diabetes duration, birth in Sweden, educational level, and baseline comorbidities [33] | 4.69 | [3.64-6.04] | Risk factor |
| Insulin use (Yes/No) | Pooled HR Adjusted for age, gender, BMI, hypertension, smoking status, and cholesterol [34] | 1.43 | [1.14-1.79] | Risk factor |
| HbA1c per 1% increase | Pooled HR Adjusted for age, gender, BMI, hypertension, smoking status, and cholesterol [34] | 1.13 | [1.12-1.15] | Risk factor |
| Fasting glucose – 1 SD increase | Pooled HR Adjusted for age, gender, BMI, hypertension, smoking status, and cholesterol [34] | 1.27 | [1.10-1.47] | Risk factor |
| MI (first-time) with covariates: |  |  |  |  |
| Age at admission (divided into 10 year bands) | HR Adjusted for gender, age, comorbidity and year of admission, invasive procedures and cardiovascular medications [21] | 1.44 | [1.43-1.45] | Risk factor |
| Sex:MaleFemale | HR Adjusted for gender, age, comorbidity and year of admission, invasive procedures and cardiovascular medications [21] | 1.000.94 | [0.93-0.96] | ReferenceRisk factor |
| Diabetes (Type I/II) | HR Adjusted for gender, age, comorbidity and year of admission, invasive procedures and cardiovascular medications [21] | 1.26 | [1.21-1.32] | Risk factor |
| Atrial Fibrillation | HR Adjusted for gender, age, comorbidity and year of admission, invasive procedures and cardiovascular medications [21] | 1.04 | [0.99-1.08] | NS |
| Cerebral Vascular Disease | HR Adjusted for gender, age, comorbidity and year of admission, invasive procedures and cardiovascular medications [21] | 1.14 | [1.11-1.17] | Risk factor |
| Congenital Heart Disease |  |  |  |  |
| Ventricular septal defect | Crude OR [25] | 1.54 | [1.31-1.81] | Risk factor |
| Tetralogy of Fallot | Crude OR [25] | 1.48 | [1.04-2.11] | Risk factor |
| Atrioventricular septal defect | Crude OR [25] | 1.84 | [1.10-3.09] | Risk factor |
| Transposition of the great arteries | Crude OR [25] | 2.31 | [1.26-4.24] | Risk factor |
| BMI (kg/m2) |  |  |  |  |
| Underweight (BMI<18.5) | HR Adjusted for years of education and body height [35] | 1.14 | [0.46-2.82] | NS |
| Normal (BMI 18.5-24.9) | HR Adjusted for years of education and body height [35] | 1.00 |  | Reference |
| Overweight (BMI 25.0-29.9) | HR Adjusted for years of education and body height [35] | 2.49 | [1.56-3.99] | Risk factor |
| Obesity (BMI ≥30) | HR Adjusted for years of education and body height [35] | 4,52 | [1.96-10.43] | Risk factor |
| Ethnicity  |  |  |  |  |
| White | Multivariable Adjusted HR [28] | 1.00 |  | Reference |
| Black | Multivariable Adjusted HR [28] | 0.46 | [0.26-0.82) | Protective factor |
| Chinese | Multivariable Adjusted HR [28] | 1.53 | [0.64-3.67] | NS |
| Hispanic | Multivariable Adjusted HR [28] | 0.66 | [0.34-1.30] | NS |
| Gender |  |  |  |  |
| Female  | Multivariable Adjusted HR [28] | 0.89 | [0.54-1.46] | NS |
| New-onset HFrEF- Females  | Stratified multivariate sub-hazard ratio [48] | 0.47 | [0.26-0.76] | Protective factor |
| New-onset HFpEF- Females  | Stratified multivariate sub-hazard ratio [48] | 2.16 | [1.21-3.83] | Risk factor |
| Alcohol Intake (drinks/week) |  |  |  |  |
| Non-drinker | HR Adjusted for age, smoking, BMI, social class, prevalent stroke, diabetes and angina [49] | 0.98 | [0.59-1.63] | NS |
| <1 | HR Adjusted for age, smoking, BMI, social class, prevalent stroke, diabetes and angina [49] | 1.43 | [0.89-2.32] | NS |
| 1-6 | HR Adjusted for age, smoking, BMI, social class, prevalent stroke, diabetes and angina [49] | 1.00 |  | Reference |
| 7-14 | HR Adjusted for age, smoking, BMI, social class, prevalent stroke, diabetes and angina [49] | 0.94 | [0.62-1.43] | NS |
| 15-34 | HR Adjusted for age, smoking, BMI, social class, prevalent stroke, diabetes and angina [49] | 1.15 | [0.78-1.70] | NS |
| ≥35 | HR Adjusted for age, smoking, BMI, social class, prevalent stroke, diabetes and angina [49] | 1.90 | [1.02-3.54] | Risk factor |
| Sleep Apnoea  |  |  |  |  |
| Non-snorers without EDS | HR Adjusted for age, waist circumference, smoking, and alcohol dependence [40] | 1.00 |  | Reference |
| Snorers without EDS | HR Adjusted for age, waist circumference, smoking, and alcohol dependence [40] | 0.77 | [0.40-1.49] | NS |
| Non-snorers with EDS | HR Adjusted for age, waist circumference, smoking, and alcohol dependence [40] | 1.33 | [0.68-2.61] | NS |
| Snorers with EDS | HR Adjusted for age, waist circumference, smoking, and alcohol dependence [40] | 2.19 | [1.14-4.18] | Risk factor |

## HF incidence and prevalence

Seferovic et al (2013) conducted a real-life contemporary analysis on the incidence and prevalence of HF in 33 countries using data from HF National Societies (HFNS) representatives in October 2011 [50]. Of the 33 countries incidence rates and/or prevalence estimates were available from 22. The countries which responded to the survey represented a population of approximately 796 million inhabitants.

Table 3: Incidence rates and prevalence of HF worldwide in studies identified by HFNS analysis [50]

| Country | Country Population | No of inhabitants ≥65 years, % | Incidence, 1000/year | Prevalence, % |
| --- | --- | --- | --- | --- |
| Belgium | 11 076 847 | 18 | NA | 2-3 |
| Czech Republic | 10 446 157 | 15 | NA | 1.5 |
| Denmark | 5 566 856 | 16 | 1.2 | 1 |
| France | 63 460 768 | 16 | 3.3 | 2.2 |
| Germany | 81 471 834 | 20 | 8.8 | NA |
| Greece | 10 787 690 | 19 | NA | 2-10 |
| Hungary | 9 905 596 | 16 | 3 | 1.6 |
| Ireland | 4 581 269 | 12 | 4 | 2 |
| Israel | 7 411 000 | 10 | NA | 1.5 |
| Italy | 59 464 644 | 20 | 1 | 3.5 |
| Lithuania | 3 349 872 | 16 | 1.41 | 2.98 |
| Norway | 4 937 305 | 15 | NA | 1-2 |
| Portugal | 10 707 924 | 17 | NA | 4.4 |
| Romania | 21 524 042 | 15 | 4 | 4.2 |
| Russia | 143 115 706 | 14 | NA | 7 |
| Serbia | 7 379 339 | 17 | 3.1 | 2.3 |
| Slovakia | 5 410 371 | 12 | NA | 2 |
| Slovenia | 2 058 032 | 16 | NA | 1-2 |
| Spain | 45 989 016 | 18 | NA | 5 (in >45 years) |
| Sweden | 9 276 509 | 18 | NA | 2 |
| Switzerland | 7 991 600 | 16 | NA | 2 |
| UK | 61 524 872 | 16 | NA | 1-2 |

The British Heart Foundation Cardiovascular Disease Statistics Report (2014) estimated the prevalence of HF (see Table 4 below) based on records from a sample of general practices across the UK using data from the Clinical Practice Research Datalink (CPRD).[51] This report only presented data on doctor-diagnosed HF. All estimates for the different age groups are age-standardised to the European Standard Population. This table is adapted from reference [51], which is based on data from the CPRD GOLD database, 2014.

Table : prevalence estimates for HF in England, Scotland, Wales, Northern Ireland and United Kingdom, 2013, by gender and age[51]

|  | England | Scotland | Wales | Northern Ireland | United Kingdom |
| --- | --- | --- | --- | --- | --- |
| **(%)** | **(%)** | **(%)** | **(%)** | **(%)** |
| Men (age/years) |  |  |  |  |  |
| 0-44 | 0.05 | 0.06 | 0.07 | 0.06 | 0.05 |
| 45-54 | 0.31 | 0.40 | 0.40 | 0.20 | 0.33 |
| 55-64 | 1.04 | 1.42 | 1.42 | 0.97 | 1.12 |
| 65-74 | 2.73 | 3.72 | 3.39 | 2.97 | 2.92 |
| 75+ | 7.64 | 8.72 | 8.25 | 8.48 | 7.84 |
| All ages | 1.17 | 1.44 | 1.37 | 1.25 | 1.22 |
| Women (age/years) |  |  |  |  |  |
| 0-44 | 0.03 | 0.06 | 0.05 | 0.02 | 0.04 |
| 45-54 | 0.15 | 0.18 | 0.18 | 0.12 | 0.15 |
| 55-64 | 0.43 | 0.55 | 0.52 | 0.54 | 0.45 |
| 65-74 | 1.22 | 1.56 | 1.80 | 1.34 | 1.32 |
| 75+ | 5.80 | 5.97 | 6.37 | 6.22 | 5.89 |
| All ages | 0.74 | 0.82 | 0.87 | 0.79 | 0.76 |
| No. of cases in sample |  |  |  |  |  |
| Men | 16,618 | 2,942 | 2,638 | 756 | 22,954 |
| Women | 13,382 | 2,094 | 2,089 | 636 | 18,201 |
| Total no. of cases in sample | 30,000 | 5,036 | 4,727 | 1,392 | 41,155 |

### Quality & Outcomes Framework Data

Given that Table 4 uses GP diagnoses from a fairly representative sample, it should show similar prevalence to data obtained from the Quality & Outcomes Framework of the GP contract. The 2014-15 prevalence in England was 0.72% and in 2015-16 it was 0.76%, somewhat lower than CPRD. However coding may be somewhat better in CPRD practices than UK practices in general.

# Methods

## HF prevalence from UK primary care and hospital data: Clinical Practice Research Datalink & Hospital Episode Statistics

### What is the Clinical Practice Research Datalink?

The Clinical Practice Research Datalink (CPRD) is an ongoing primary care database of longitudinal anonymised electronic health records (EHRs) from general practitioners, with coverage of over 11.3 million patients from 674 practices in the UK. With 4.4 million active (alive, currently registered) patients meeting quality criteria, approximately 6.9% of the UK population are included and patients are broadly representative of the UK general population in terms of age, sex and ethnicity. The distribution of CPRD practices is shown in Figure 1 below.

Figure : Distribution of the 674 CPRD practices by region in England, Wales, Scotland and Northern Ireland



### Identification of HF cases and code lists for CPRD and HES doctor diagnosed cases

Table 5 contains the list of Medcodes and Read Codes that were used to identify the CPRD doctor diagnosed cases, and Table 6 shows the ICD-10 codes that were used for the identification of HES doctor diagnosed HF cases. Data was extracted from CPRD (<http://www.cprd.com/intro.asp>) to analyse the prevalence of HF. HF cases were identified in four ways:

* CPRD doctor diagnosed cases
* Hospital Episode Statistics (HES) inpatient doctor diagnosed cases
* cases that were inferred using records of symptoms, signs and test results that were not explicitly doctor diagnosed, using a clinical algorithm based on the most recent NICE and ESC guidelines, and
* patients who had been prescribed HF drugs and had clinical HF signs but were not doctor diagnosed, using a drugs algorithm.

A comprehensive list of Medcodes, Prodcodes and ICD-10 codes was compiled for the identification of the CPRD and HES doctor diagnosed HF cases, and for the undiagnosed (but diagnosable) clinical and drugs algorithm HF cases. The clinical and drugs algorithms were developed based on the criteria found in the NICE and European Society of Cardiology (ESC) Guidelines.[6 ,52] The Medcodes used for CPRD HF diagnoses are shown in Table 5. These include all QOF HF diagnostic Read codes, plus other codes which reliably indicate a previous diagnosis e.g. “preferred place of care for next exacerbation heart failure”.

Table : Medcodes that were used for the identification of CPRD doctor diagnosed HF cases

| Read Term | Medcode | Read Code | ICD-10 Code |
| --- | --- | --- | --- |
| H/O: heart failure | 15058 | 14A6.00 | I50.9 |
| H/O: heart failure in last year | 46912 | 14AM.00 | I50.9 |
| Heart failure confirmed  | 9913 | 1O1..00 | I50.9 |
| Echocardiogram shows left ventricular systolic dysfunction | 11284 | 585f.00 | I50.1 |
| Echocardiogram shows left ventricular diastolic dysfunction | 11351 | 585g.00 | I50.1 |
| Heart failure self-management plan agreed | 106198 | 661M500 | I50.9 |
| Congestive heart failure monitoring | 12366 | 662T.00 | I50.0 |
| Heart failure annual review | 30779 | 662W.00 | I50.9 |
| New York Heart Association classification - class I | 18853 | 662f.00 | I50.9 |
| New York Heart Association classification - class II | 13189 | 662g.00 | I50.9 |
| New York Heart Association classification - class III | 19066 | 662h.00 | I50.9 |
| New York Heart Association classification - class IV | 51214 | 662i.00 | I50.9 |
| Heart failure 6 month review | 83502 | 662p.00 | I50.9 |
| Education about deteriorating heart failure | 105002 | 679W100 | I50.9 |
| Cardiac failure therapy | 24503 | 8B29.00 | I50.9 |
| Heart failure care plan discussed with patient | 32945 | 8CL3.00 | I50.9 |
| Has heart failure management plan | 103732 | 8CMK.00 | I50.9 |
| Heart failure clinical pathway | 106008 | 8CMW800 | I50.9 |
| Preferred place of care for next exacerbation heart failure | 105542 | 8CeC.00 | I50.9 |
| Admit heart failure emergency | 32898 | 8H2S.00 | I50.9 |
| Heart failure follow-up | 17851 | 8HBE.00 | I50.9 |
| Referral to heart failure exercise programme | 70619 | 8HHz.00 | I50.9 |
| Seen in heart failure clinic | 12627 | 9N0k.00 | I50.9 |
| Seen by community heart failure nurse | 19002 | 9N2p.00 | I50.9 |
| Referred by heart failure nurse specialist | 69062 | 9N6T.00 | I50.9 |
| Malignant hypertensive heart disease with CCF | 72668 | G210100 | I11.0 |
| Benign hypertensive heart disease with CCF | 52127 | G211100 | I11.0 |
| Hypertensive heart disease NOS with CCF | 62718 | G21z100 | I11.0 |
| Hypertensive heart & renal disease with (congestive) heart failure | 21837 | G232.00 | I13.0 |
| Hypertensive heart & renal disease + both (congestive) heart and renal failure | 57987 | G234.00 | I13.2 |
| Heart failure  | 2062 | G58..00 | I50.9 |
| Cardiac failure | 1223 | G58..11 | I50.9 |
| Congestive heart failure | 398 | G580.00 | I50.0 |
| Congestive cardiac failure | 2906 | G580.11 | I50.0 |
| Right heart failure | 10079 | G580.12 | I50.0 |
| Right ventricular failure | 10154 | G580.13 | I50.0 |
| Biventricular failure | 9524 | G580.14 | I50.9 |
| Acute congestive heart failure | 23707 | G580000 | I50.0 |
| Chronic congestive heart failure | 32671 | G580100 | I50.0 |
| Decompensated cardiac failure | 27884 | G580200 | I50.9 |
| Compensated cardiac failure | 11424 | G580300 | I50.9 |
| Congestive heart failure due to valvular disease | 94870 | G580400 | I50.0 |
| Left ventricular failure | 884 | G581.00 | I50.1 |
| Impaired left ventricular function | 5942 | G581.13 | I50.1 |
| Acute left ventricular failure | 5255 | G581000 | I50.1 |
| Acute heart failure | 27964 | G582.00 | I50.9 |
| Heart failure with normal ejection fraction | 101138 | G583.00 | I50.9 |
| HFNEF – heart failure with normal ejection fraction | 101137 | G583.11 | I50.9 |
| Heart failure with preserved ejection fraction | 106897 | G583.12 | I50.9 |
| Right ventricular failure | 104275 | G584.00 | I50.0 |
| Heart failure NOS | 4024 | G58z.00 | I50.9 |
| Cardiac failure NOS | 17278 | G58z.12 | I50.9 |
| Post cardiac operation heart failure NOS | 96799 | G5y4z00 | I50.9 |
| Left ventricular systolic dysfunction | 8966 | G5yy900 | I50.1 |
| Left ventricular diastolic dysfunction | 12550 | G5yyA00 | I50.1 |
| Heart failure as a complication of care | 66306 | SP11111 | I50.9 |

Table 6 shows the ICD-10 codes that were used for the identification of HES doctor diagnosed HF cases.

Table : ICD-10 codes that were used for the identification of HES doctor diagnosed HF cases

|  |  |
| --- | --- |
| ICD-10 Code | Diagnosis |
| I50.0 | Congestive Heart Failure |
| I50.1 | Left ventricular failure |
| I50.9 | Heart failure, unspecified  |
| I11.0 | Hypertensive heart disease with (congestive) heart failure |
| I13.0 | Hypertensive heart and renal disease with (congestive) heart failure |
| I13.2 | Hypertensive heart and renal disease with both (congestive) heart failure and renal failure |

### Additional HF cases from clinical algorithm

In order to identify additional undiagnosed (but diagnosable) HF cases, the NICE Acute and Chronic HF Guidelines and the 2016 ESC HF Guidelines were used to develop two algorithms– a clinical algorithm and a drugs algorithm. The clinical algorithm combines information about specific symptoms and signs of HF with data from natriuretic peptide test results and ECG and echocardiogram results in order to identify additional probable HF cases. The flowchart of the clinical algorithm that was developed is shown below in Figure 2: clinical algorithm flowchart.

Figure : clinical algorithm flowchart

Develop the Clinical algorithm

Keep cases who have not had echocardiogram

(N=162,731)

Keep abnormal echocardiogram cases

clinical algorithm – keep cases fulfilling criteria

History of myocardial infarction, orthopnoea, paroxysmal nocturnal dyspnoea

OR

Signs of elevated JVP, displaced apex beat

OR

Abnormal ECG

Exclude HF cases with HES or CPRD diagnosis

AND

NT-proBNP ≥125 pg/mL

Or

BNP ≥35 pg/mL

Additional abnormal echocardiogram probable HF cases

Clinical algorithm probable HF cases

Exclude HF cases with HES or CPRD diagnosis

Additional clinical algorithm probable HF cases

Probable HF cases

According to the 2016 ESC Guidelines,[6] echocardiography is the single most valuable test in diagnosing HF because structural abnormality, systolic dysfunction, diastolic dysfunction, or a combination of these factors are required to be recognised in patients who have resting and/or exertional symptoms of HF, to conclude a definitive HF diagnosis. Therefore, the first step of our clinical algorithm was to identify patients who had received an abnormal echocardiogram result, and after excluding doctor-diagnosed cases these patients were labelled as probable HF cases.

The next step involved the identification of CPRD patients who had displayed the typical symptoms or specific signs of HF, or abnormal ECG results. The ESC Guidelines state that HF is unlikely in patients presenting with a completely normal ECG. If the identified patients also had N-terminal pro-B-type natriuretic peptide (NTproBNP) levels of ≥125pg/mL or BNP levels of ≥35pg/mL then they were labelled as probable clinical algorithm cases. B-type natriuretic peptide (BNP) is a neurohormone that is secreted mostly in the cardiac ventricles due to pressure overload and volume expansion [53]. Plasma concentration of natriuretic peptides can be used as an initial diagnostic HF test, especially when echocardiography is unavailable. Patients who have natriuretic peptide levels below the cut off points can be ruled out as HF patients. The 2016 ESC Guidelines stated that NTproBNP levels of ≥125pg/mL or BNP levels of ≥35pg/mL could be used as an indication of HF.

### Additional HF cases from drugs algorithm

Beta blockers have been found to reduce morbidity and mortality in patients with HF with reduced left ventricular ejection fraction.[54] However, present guidelines make no differentiation in treatment based on age or sex. There is a greater likelihood of patients being female, with increasing age. In many RCTs, there is a moderately low number older patients, which has resulted in an uncertainty regarding the optimum management of HFrEF in older patients.[54] Kotecha et al (2016) found that beta-blockers reduced mortality hazard rations significantly and the number of hospital admissions associated with HF in all sex and age subgroups.[55]

Drug prescriptions are an additional potential source of diagnostic data. NICE recommends giving both angiotensin-converting enzyme inhibitors (ACEIs) and beta-blockers (BBs) to all patients with heart failure, and in those patients who remain symptomatic an aldosterone antagonist, or an angiotensin II receptor antagonist (ARB), or hydralazine in combination with nitrate if the patient is of African or Caribbean origin. However, we will exclude patients on these drugs with a diagnosis of hypertension of coronary heart disease (CHD), and will explore methods of discriminating heart failure based on NICE’s recommended prescribing pattern. At this stage it appears that a combination of ACEI/ARB + BB licensed for HF (carvedilol, bisoprolol or nebivolol) is likely to be systolic heart failure, and this is even more likely if spironolactone or eplerenone are also prescribed. It is difficult to identify a group that would have these three classes of drug that would not have systolic heart failure.

The drugs algorithm we developed combined drug information with HF signs and symptoms and abnormal ECG results, in order to identify possible HF cases which were not recorded in HES or CPRD as doctor-diagnosed cases**.** The flowchart of the drugs algorithm that was developed is shown below in Figure 3.

**Figure 3: drugs algorithm Flowchart**

Develop the Drugs algorithm

Exclude Hypertension and CHD cases

Drugs algorithm probable HF cases

Exclude HF cases with HES or CPRD or clinical algorithm diagnosis

Additional drugs algorithm probable HF cases

drugs algorithm – keep cases fulfilling criteria

History of myocardial infarction, orthopnoea, paroxysmal nocturnal dyspnoea

OR

Signs of elevated JVP, displaced apex beat

OR

Abnormal ECG

AND

Patients prescribed ACEi/Angiotensin II receptor antagonists/Beta-blocker/potassium sparing diuretics in CPRD

Probable HF cases

### CPRD risk factors

We used the literature review described in the Background to extract CPRD 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 CPRD. For example, data on educational level, occupational class and socioeconomic status is very poorly recorded. The occupational classification for which Read codes are available is from a 1986 Office for National Statistics classification so is outdated. CPRD links most patients’ data to Index of Multiple Deprivation (IMD) data based on postcode. Secondly, to produce local estimates we use “joint distributions”- cross tabulations which distribute data on each risk factor across the data for all other risk factors- of local risk factor data to which we apply the CPRD prevalence estimates for the same distributions. Hence we can only use in the final regression model variables which are also available locally. This may cause model performance to deteriorate. We evaluated the extent of this by comparing Receiver Operating Characteristic (ROC) curves for the two models. Although, the literature review highlights a significant association between congenital heart disease and HF, congenital heart disease was not included as a risk factor in the final model. This is mainly due to the prevalence of congenital heart in the UK being less than 1% during 2000 to 2005.[47] Therefore, as the prevalence is low, congenital heart disease may not be a significant risk factor for HF in the UK.

Risk factor data were extracted by a defined Read code lists, and by extracting additional data using the ‘adid’, ‘enttype’ and ‘data’ variables in the CPRD Additional datasets. Read code lists were created by searching for relevant Read version 2 5-byte codes using either CPRD’s own code browser or using the “NHS browser” maintained by the Health & Social Care information Centre (HSCIC). We used the NHS browser and CPRD browser to create code lists for ethnicity, BMI, smoking, alcohol consumption, CHD, hypertension, diabetes, sleep apnoea and atrial fibrillation by searching relevant read terms or and selecting the appropriate medcodes. The smoking, alcohol and BMI data extracted using medcodes was combined with information found in the CPRD additional datasets, and the most recent recording for each patient was selected using the ‘eventdate’ variable.

### CPRD HF Control group

Due to time constraints, a 1:10 random sample was taken from the CPRD cohort for the control group, rather than using the entire CPRD cohort. The sample was extracted with approximately 1.05 million CPRD patients. Any patients below the age of 16 were excluded from the sample with only those 16 years of age and above remaining in the control group. Patients above the age of 105 years were also excluded.

### CPRD descriptive analyses

We performed a number of descriptive analyses on the patient-level dataset including demographics, risk factor breakdowns and categories.

### CPRD prevalence and incidence

The overall and age and gender-specific prevalence and incidence of HF was estimated using denominator data from CPRD. (For sensitivity analysis, the primary-care data from CPRD was combined with secondary-care data from the linked HES data). Patients were identified by Read code and the most suitable ICD-10 classification. Period prevalence was calculated for 2016 using the number of patients that are alive and registered within an up-to-standard (UTS) practice on the beginning of the year point (January 2016) as the denominator. Annual incidence was estimated with regards to the average annual incidence of HF observed since at the beginning of the year.

### 1.1.1 CPRD regression modelling

We fitted uni-variate then multivariate logistic regression models for HF as described in previous publications, to produce odds ratios (ORs) and regression coefficients.[56] A range of multivariate regression models were fitted in order to obtain the best performing. We included one additional variable at a time to observe the effects.

### Interactions

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.[57] For example, there might be an interaction between the HF risk factors of education level and social class. An alternative term for interaction ifs effect modification. In this example, we can think of this as educational level modifying the effect of social class. The most flexible approach to examine interactions is to use regression models, but when using Mantel-Haenszel methods to control for confounding an alternative is to use a χ2 test for effect modification, commonly called a test of heterogeneity. Interaction, effect modification and heterogeneity are three different ways of describing the same thing. Log likelihoods are compared in the two models excluding and including the interaction parameters to test the null hypothesis that there is no interaction between selected variables.

We tested for interactions between CPRD predictor variables for risk factors. Interactions were tested between age and gender; BMI and smoking status; BMI and economic activity; economic activity and education; age and education; age and socioeconomic status; age and smoking status; age and economic status; age and BMI.

### Internal validation

We fitted a range of multivariate logistic regression models for in order to obtain the best performing. We included one additional variable at a time to observe the effects. In order to obtain the most parsimonious models we then applied stepwise backward and forward variable selection using the *stepwise* command in Stata. 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 CPRD patient the probability of having HF using the derived odds ratios (ORs), and by using these probabilities to examine sensitivity and specificity.

The variables included in the final model are also determined by the availability of local data to match with the model variables. Hence variable selection has to be a compromise between the best model which can be produced from CPRD data and the local variable available.

All statistical analysis was carried out in StataSE14.

### Local prevalence estimates

Derived ORs (or 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, smoking, ethnicity, cardiovascular diseases and other disease conditions. The local model uses locally available data.

The “local” model includes only those variables that are available at local population level i.e. age, sex, socioeconomic status, BMI, smoking status, depression 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/smoking 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.

We have developed two methods for producing small population estimates and associated CIs in Stata software. One uses a bootstrapping method to produce repeated samples (Method 1), the other (Method 2) uses inverse probability weights. Both methods produce CIs for the estimates, which are derived from the variance in the logistic model, not the local populations. It would have been useful to compare the results of both methods, but because of the short timeframe for this project we only used Method 2: Logistic regression and inverse probability weights.

* + 1. **Method 1: bootstrapping procedure to produce repeated samples**

The detailed methods of the Stata code we developed and used is included in Annex 1: synthetic estimation using Stata. 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 as follows, which results in one line of data per practice or MLSOA (using the population identifier as the by variable) in Stata.

We 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 CPRD 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).

* + 1. **Method 2: Logistic regression and inverse probability weights**

Inverse probability weighting methods are used to standardise 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.

In a population case-control study, our sampled population is an exhaustive list of disease cases, plus a random sample of controls without the disease, with a known sampling fraction. The sampling probability weights are inversely proportional to the sampling fraction for each sub-population. For cases, the sampling probability weight is 1. And, for controls, the sampling probability weight is the reciprocal of the sampling fraction. (So, if the sampling fraction is 1/8, then controls are weighted upwards by a factor of 8.) These sampling-probability weights are used in logistic regression models. Predicted disease probabilities from these models will then be unbiased, if the model is correctly specified.

Similarly to Method 1 we estimated population parameters for logistic regression models. The risk factors in the model fell into two classes, namely always-present risk factors and sometimes-missing risk factors. The always-present risk factors were gender (Male or Female), age group (18-44, 45-64, 65-74 and 75+), ethnicity (White, Mixed, Black, Asian or Other, imputed to White if not known). The sometimes-missing risk factors were practice index of multiple deprivation (IMD) quintile (1, 2, 3, 4 or 5), smoking status (Non-smoker, Ex-smoker or Smoker), alcohol units per week category (None, (0,14], (14,42] or >42), and body mass index in kilos/square metre (BMI) category ((0, 18.5], (18.5,25], (25,30] or >30).

We fitted the logistic regression model, using Huber variances and sampling-probability weights. The parameters were a baseline odds for each of the 2x4=8 combinations of gender and age group, an odds ratio for each ethnicity except White, an odds ratio for each IMD quintile except the first, an odds ratio for each smoking status except Non-smoker, an odds ratio for each alcohol consumption category except Zero units, and an odds ratio for each BMI category except (18.5,25] kilos per square metre. The sampling-probability weights used were equal to the products of two sets of component sampling-probability weights. The first set of component weights standardised by case status from the case-control study sample to the denominator population from which the cases and controls were sampled, and were equal to 1 for HF cases (assumed to be sampled exhaustively from the cases in the CPRD denominator population), and equal in the controls to the reciprocal of the sampling fraction of the controls as a fraction of the non-cases in the CPRD denominator population (equal to 27.211693).

We also use inverse probability weights to correct for missing values as an easy-to-use alternative to multiple imputation. We then define the inverse probability weights using a completeness-propensity score. We have a panel of variables *V1…VK* that are always present (such as age and gender), and a panel of variables *U1…UJ* that are sometimes missing. Let C (for completeness) be the binary indicator variable indicating that all the variables *U1…UJ* are present. We then use a logistic regression model, regressing C with respect to the always-complete variables *V1…VK*. The completeness-propensity score is defined as the predicted completeness probability for each individual, under that regression model. The inverse-probability weight, for each individual with a complete set of data *U1…UJ*, is then the reciprocal of that individual’s completeness-propensity score. Therefore, individuals with a high probability of having complete data (like elderly females) are weighted downwards. And individuals with a low probability of completeness (like young males) are weighted upwards. These inverse-probability weights can then be used in further regression models, such as a logistic regression model to predict disease.

Therefore, the second set of component weights were computed to standardise the sample of cases and controls with all risk factors present to the total sample of cases and controls by gender, age group and ethnicity, and were derived as inverse probabilities of presence of the full set of risk factors (completeness) from a logistic regression model with completeness as the outcome, fitted to the cases and controls, using the first set of sampling-probability weights to standardise by case status, and whose parameters were a baseline odds for each of the 8 combinations of gender and age group and an odds ratio for each non-white ethnic category. The product weights therefore were computed to standardise the odds and odds ratios from the sample of cases and controls with all risk factors present (272,369 subjects out of a total of 101,870 cases and 440,293 sampled controls) to the total denominator population of subjects aged at or above 18 years, with or without HF, on their birthdays in 2015 (13,864,783 subjects). We also fitted logistic regression models of HF status with respect to the 8 combinations of gender and age only, using only the first set of sampling probability weights to standardise by HF status, in order to estimate odds (and thereby prevalence) of HF for each combination of gender and age group in the CPRD population at large.

Having estimated the regression model parameters, we used these for out-of-sample prediction of HF prevalence, using the *margprev* add-on Stata package [58 ,59]. These predicted prevalence estimates were for the sub-populations of patients for 7,692 practices, for 204 clinical care groups (CCGs), and for 6,755 MSOAs, for which information was available on the marginal frequencies of the seven risk factors in the model. We computed estimated prevalence assuming that, within each sub-population, the seven risk factors were mutually statistically independent, implying that we could give each possible combination of the seven risk factors a sampling-probability weight proportional to the product of the proportions of subjects with each of the appropriate risk-factor values. Therefore, for each subpopulation, we had 2x4x5x5x3x4x4=9600 combinations of risk factor values, with proportions of subjects calculated assuming statistical independence, and estimated the expected subpopulation prevalence of HF accordingly. The assumption of statistical independence of risk factors is probably not literally true, but might be expected to give prevalence estimates that are not vastly in error if the effects of the risk factors are not too non-additive. We have not internally or externally validated this method yet.

We have used method 2, logistic regression and inverse probability weights for other models because of the large number of variables in most of the models. This required us to produce Stata datasets of local risk factor data which have one observation for every permutation of all the risk factors for every practice, which generated very large files (up to 60 GB). We were able to process these using Stata/MP, the fastest and largest version of Stata. On dual-core chips, Stata/MP runs 40% faster overall and 72% faster on time-consuming estimation commands. It can handle a maximum number of 32,767 variables and 20 billion observations. Some of the datasets we used included over one billion observations. Processing was carried out on a multicore server. It would not have been possible to run the bootstrapping procedure to produce repeated samples which requires fitting a logistic model 1,000 times for each practice.

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 used both methods as a means of validating the synthetic estimation step. The detailed methods of the Stata code we developed and used is included in Annex 1: synthetic estimation using Stata. 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 as follows, which results in one line of data per practice or MLSOA (using the population identifier as the by variable) in Stata.

## Validation of local estimates

### Internal validation

In addition to the internal and external validation of the regression models, local estimates can also be validated by aggregating them to the lowest geography available in the raw data and comparing them, a form of internal validation. We have over time increased the number of variables used in the local models as more local data has become available. However as more variables are added we need to take account of the joint effects of multiple risk factors, i.e. it assumes they operate independently. Estimation of the joint effects of multiple risk factors is complex for several reasons. In particular, some of the effects of more distal risk factors are mediated through intermediate factors. We have acknowledged this by creating specific joint distributions for variables where this is known e.g. age and educational level, as older age groups are less likely to have tertiary education.

### External validation

There are Quality & Outcomes Framework (QOF) disease registers[60] for all the prevalence models we produced for PHE. We have experience in comparing QOF-registered prevalence and estimated prevalence right down to practice level using spatial analyses.[61] Local estimates can also be validated against the corresponding QOF register for each geography using Bland-Altman plots. This method uses graphical methods to investigate the assumptions of the method and also gives confidence intervals.[62] It aims to quantify the agreement between and clinical importance of two methods of clinical measurement using the differences between observations made using the two methods on the same subjects. The 95% limits of agreement, estimated by mean difference 1.96 standard deviation of the differences, provide an interval within which 95% of differences between measurements by the two methods are expected to lie. The second method is based on errors-in-variables regression in a classical (X,Y) plot and focuses on confidence intervals, whereby two methods are considered equivalent when providing similar measures notwithstanding the random measurement errors.[63] A recent update reconciles these two methodologies and shows their similarities and differences using both real data and simulations.[64]

The flowchart in Figure 2 shows the four methods we used to identify HF diagnoses and likely/probable HF diagnoses based on clinical and prescribing data.

**Figure 4: Flow chart for analysis of HF prevalence (Final Numbers)**

Medcode extraction of records from CPRD database

(N=507,429,744)

Number of patients

(N=224,265)

Number of patients

(N=1,794,085)

Prodcode extraction of records from CPRD database

(N=977,922,246)

Keep cases who have not had echocardiogram

(N=162,731)

Linked HES data (N=78,565)

Extract ICD-10 codes relevant to the diagnosis of HF

Exclude HF cases with HES or CPRD diagnosis (N=8,251)

Keep abnormal echocardiogram cases

(N=8,251)

drugs algorithm – keep cases fulfilling criteria

Develop the Drugs algorithm

History of myocardial infarction, orthopnoea, paroxysmal nocturnal dyspnoea

OR

Signs of elevated JVP, displaced apex beat

OR

Abnormal ECG

AND

Patients prescribed ACEi/Angiotensin II receptor antagonists/Beta-blocker/potassium sparing diuretics in CPRD

Exclude Hypertension and CHD cases (N=165,329)

Drugs algorithm probable HF cases (N=44,209)

Exclude HF cases with HES or CPRD or clinical algorithm diagnosis (N=11,363)

Additional drugs algorithm probable HF cases (N=32,846)

Identify doctor diagnosed CPRD HF cases

(N=244,235)

Exclude HF cases who have a HES diagnosis

(N=62,595)

Additional CPRD diagnosed HF cases

(N=181,640)

Additional abnormal echocardiogram probable HF cases (N=0)

Additional clinical algorithm probable HF cases (N=0)

Exclude HF cases with HES or CPRD diagnosis (N=10,249)

Extract medcodes relevant to the diagnosis of HF

Identify doctor diagnosed HES AND CPRD HF cases

(N=62,595)

Identify doctor diagnosed HES HF cases

(N=76,692)

Exclude HF cases who have a CPRD HF diagnosis

(N=62,595)

Additional HES diagnosed HF cases

(N=14,097)

Clinical algorithm probable HF cases (N=10,249)

History of myocardial infarction, orthopnoea, paroxysmal nocturnal dyspnoea

OR

Signs of elevated JVP, displaced apex beat

OR

Abnormal ECG

AND

NT-proBNP ≥125 pg/mL

Or

BNP ≥35 pg/mL

clinical algorithm – keep cases fulfilling criteria

Probable HF cases (N=291,178)

Develop the Clinical algorithm

# Results

## HF prevalence from CPRD

### Baseline characteristics of CPRD respondents

Table 7 compares the baseline characteristics of HF cases and controls. Cases are older than the control sample, and, partly because of this, more likely to have co-morbidities.

Table . Characteristics of CPRD respondents

|  | Controls | HF cases | Total |
| --- | --- | --- | --- |
| Total numbers | 1,045,427 | 288,305 | 1,333,732 |
| Age group |  |  |  |
| 16-44 | 458,019 (43.81%) | 3,388 (1.18%) | 461,407 (34.60%) |
| 45-64 | 332,545 (31.81%) | 33,446 (11.60%) | 365,991 (27.44%) |
| 65-74 | 110,040 (10.53%) | 53,656 (18.61%) | 163,696 (12.27%) |
| Over 75 | 144,823 (13.85%) | 197,815 (68.61%) | 342,638 (25.69%) |
| Gender  |  |  |  |
| Male  | 476,605 (45.59%) | 145,533 (50.48%) | 622,138 (46.65%) |
| Female  | 568,822 (54.41%) | 142,772 (49.52%) | 711,594 (53.35%) |
| Ethnicity |  |  |  |
| White | 897,344(85.84%) | 253,016(87.76%) | 1,150,360(86.25%) |
| Black  | 13,380(1.28%) | 1,417(0.49%) | 14,797(1.11%) |
| Asian | 25,561(2.45%) | 2,807(0.97%) | 28,368(2.13%) |
| Mixed  | 102,945(9.85%) | 30,502(10.58%) | 133,447(10.01%) |
| Other  | 6,197(0.59%) | 563(0.20%) | 6,760(0.51%) |
| BMI |  |  |  |
| <18.4 underweight | 51,347 (6.32%) | 8,766 (3.92%) | 60,113 (5.80%) |
| 18.5 – 24 normal | 371,472 (45.69%) | 71,774 (32.12%) | 443,246 (42.77%) |
| 25 – 29 overweight | 240,715 (29.61%) | 75,084 (33.60%) | 315,799 (30.47%) |
| >30 obese  | 149,472 (18.39%) | 67,823 (30.35%) | 217,295 (20.97%) |
| Smoking |  |  |  |
| Non- smoker | 522,810 (56.38%) | 131,413 (48.47%) | 654,223 (54.59%) |
| Ex-smoker | 162,203 (17.49%) | 101,177 (37.32%) | 263,380 (21.98%) |
| Current smoker | 242,333 (48.47%) | 38,519 (14.21%) | 280,852 (23.43%) |
| CHD (excluding MI) |  |  |  |
| No  | 1,018,062 (97.38%) | 220,296 (76.41%) | 1,238,358 (92.85%) |
| Yes | 27,365 (2.62%) | 68,009 (23.59%) | 95,374 (7.15%) |
| Hypertension |  |  |  |
| No  | 901,794 (86.26%) | 141,983 (49.25%) | 1,043,777 (78.26%) |
| Yes | 143,633 (13.74%) | 146,322 (50.75%) | 289,955 (21.74%) |
| Diabetes |  |  |  |
| No | 929,129 (88.88%) | 215,593 (74.78%) | 1,144,722 (85.83%) |
| Type 1 or Type II | 116,298 (11.12%) | 72,712 (25.22%) | 189,010 (14.17%) |
| Atrial Fibrillation |  |  |  |
| No  | 1,027,363 (98.27%) | 209,999 (72.84%) | 1,237,361 (92.77%) |
| Yes | 18,064 (1.73%) | 78,306 (27.16%) | 96,370 (7.23%) |
| Alcohol Intake (Units/Week) |  |  |  |
| Non-drinker | 99,349 (14.90%) | 44,842 (20.27%) | 144,191 (16.23%) |
| (0, 14] | 387,247 (58.06%) | 121,448 (54.91%) | 508,695 (57.27%) |
| (14, 42] | 156,602 (23.48%) | 48,199 (21.79%) | 204,801 (23.06%) |
| >42 | 23,791 (3.57%) | 6,687 (3.02%) | 30,478 (3.43%) |

### 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, where recording in primary care is still low, followed by BMI. With these two exceptions, all other variables missed 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  | 2014 |
|  | Frequency | %\* |
| BMI  | 297,279 | 22.29% |
| Smoking | 135,277 | 10.14% |
| Alcohol Consumption Status  | 445,567 | 33.41% |

### CPRD prevalence and incidence

Prevalence and incidence of HF in the CPRD data were calculated for doctor-diagnosed and for algorithm-diagnosed HF (or high risk of HF, not shown here) using the two diagnostic and two algorithm-based methods. Prevalence was inferred from cumulative incidence, with HF cases removed only at death.

The prevalence of HF for the years 1959-2015 is shown in Table 9. The prevalence of HF is estimated at around 0.74% for doctor-diagnosed HF, rising if algorithm cases are included. The incidence of HF for the years 2000-2015 is also shown in Table 9.

Table . Prevalence and incidence of doctor-diagnosed HF in the CPRD data: 1959-2015.

| Year | Prevalence of doctor diagnosed HF (cases per million) | Incidence of doctor diagnosed HF (cases per million) |
| --- | --- | --- |
| 1959 | 43.03 | 2.10 |
| 1960 | 45.89 | 4.35 |
| 1961 | 49.15 | 4.82 |
| 1962 | 50.94 | 3.69 |
| 1963 | 54.32 | 5.29 |
| 1964 | 57.98 | 5.62 |
| 1965 | 67.15 | 11.33 |
| 1966 | 72.45 | 7.53 |
| 1967 | 78.67 | 8.39 |
| 1968 | 86.39 | 10.04 |
| 1969 | 100.29 | 16.36 |
| 1970 | 123.73 | 26.44 |
| 1971 | 137.60 | 17.71 |
| 1972 | 156.41 | 23.04 |
| 1973 | 183.17 | 31.65 |
| 1974 | 215.15 | 37.47 |
| 1975 | 243.72 | 35.04 |
| 1976 | 276.67 | 40.97 |
| 1977 | 316.69 | 49.24 |
| 1978 | 365.79 | 59.40 |
| 1979 | 420.87 | 67.56 |
| 1980 | 507.80 | 102.03 |
| 1981 | 576.54 | 86.78 |
| 1982 | 675.47 | 119.05 |
| 1983 | 777.14 | 124.78 |
| 1984 | 897.06 | 145.96 |
| 1985 | 1065.64 | 198.48 |
| 1986 | 1258.59 | 228.79 |
| 1987 | 1475.51 | 260.14 |
| 1988 | 1830.44 | 411.14 |
| 1989 | 2362.92 | 644.34 |
| 1990 | 3127.21 | 988.47 |
| 1991 | 3781.12 | 1014.06 |
| 1992 | 4254.40 | 919.51 |
| 1993 | 4726.09 | 963.52 |
| 1994 | 5241.43 | 1033.63 |
| 1995 | 5595.01 | 926.04 |
| 1996 | 5936.69 | 908.92 |
| 1997 | 6175.78 | 863.12 |
| 1998 | 6396.61 | 885.22 |
| 1999 | 6627.69 | 892.07 |
| 2000 | 6864.90 | 925.63 |
| 2001 | 7091.87 | 910.34 |
| 2002 | 7251.80 | 878.02 |
| 2003 | 7346.56 | 809.41 |
| 2004 | 7378.48 | 726.31 |
| 2005 | 7343.19 | 639.74 |
| 2006 | 7272.31 | 561.74 |
| 2007 | 7194.47 | 524.00 |
| 2008 | 7154.91 | 520.99 |
| 2009 | 7175.63 | 523.40 |
| 2010 | 7210.58 | 512.24 |
| 2011 | 7258.00 | 498.25 |
| 2012 | 7305.28 | 495.74 |
| 2013 | 7358.98 | 460.46 |
| 2014 | 7409.44 | 417.35 |
| 2015 | 7444.42 | 353.95 |

Table 10 shows the prevalence rates of doctor-diagnosed HF 2000-2015 in males by age at diagnosis, and Table 11 shows the corresponding data for females. The prevalence and incidence rates of algorithm cases (not shown here) as a percentage of doctor diagnosed cases decreases with increasing age. This could be consistent with the algorithm- diagnosed cases being an at-risk or undiagnosed group who are later confirmed as having HF.

Table : prevalence rates of doctor-diagnosed HF 2000-2015 in males by age at diagnosis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 16-44 | 45-64 | 65-74 | 75+ | All ages |
| 2000 | 694.5 | 5068.5 | 8775.5 | 18200.0 | 32738.5 |
| 2001 | 740.0 | 5266.0 | 9015.5 | 19618.5 | 34640.0 |
| 2002 | 801.5 | 5468.5 | 9314.0 | 20709.0 | 36293.0 |
| 2003 | 850.5 | 5700.0 | 9416.0 | 21701.5 | 37668.0 |
| 2004 | 924.0 | 5943.5 | 9532.0 | 22430.5 | 38830.0 |
| 2005 | 991.5 | 6063.0 | 9421.5 | 23120.0 | 39596.0 |
| 2006 | 1040.5 | 6170.5 | 9366.5 | 23660.5 | 40238.0 |
| 2007 | 1088.0 | 6266.0 | 9251.5 | 24116.0 | 40721.5 |
| 2008 | 1129.0 | 6358.5 | 9186.5 | 24713.0 | 41387.0 |
| 2009 | 1155.0 | 6374.0 | 9162.5 | 25631.5 | 42323.0 |
| 2010 | 1177.0 | 6474.0 | 9223.0 | 26442.5 | 43316.5 |
| 2011 | 1193.0 | 6542.0 | 9296.5 | 27310.0 | 44341.5 |
| 2012 | 1216.5 | 6563.5 | 9362.5 | 28231.5 | 45374.0 |
| 2013 | 1226.0 | 6509.5 | 9417.5 | 29299.5 | 46452.5 |
| 2014 | 1200.0 | 6459.0 | 9427.5 | 30263.0 | 47349.5 |
| 2015 | 1173.5 | 6375.5 | 9371.5 | 31300.0 | 48220.5 |

Table : prevalence rates of doctor-diagnosed HF 2000-2015 in females by age at diagnosis

| Year | 16-44 | 45-64 | 65-74 | 75+ | All ages |
| --- | --- | --- | --- | --- | --- |
| 2000 | 674.0 | 2584.5 | 6917.5 | 32652 | 42828.0 |
| 2001 | 709.5 | 2633.0 | 6857.5 | 34458.0 | 44658.0 |
| 2002 | 735.5 | 2634.5 | 6839.0 | 35801.5 | 46010.5 |
| 2003 | 759.5 | 2651.0 | 6674.5 | 36826.0 | 46911.0 |
| 2004 | 810.0 | 2665.0 | 6405.5 | 37426.0 | 47306.5 |
| 2005 | 821.5 | 2600.5 | 6078.5 | 37783.0 | 47283.5 |
| 2006 | 842.0 | 2639.0 | 5765.0 | 37675.0 | 46921.0 |
| 2007 | 857.0 | 2619.5 | 5513.0 | 37584.0 | 46573.5 |
| 2008 | 868.0 | 2608.0 | 5321.5 | 37627.0 | 46424.5 |
| 2009 | 898.0 | 2620.0 | 5158.0 | 38017.0 | 46693.0 |
| 2010 | 919.0 | 2662.5 | 5005.0 | 38447.5 | 47034.0 |
| 2011 | 930.0 | 2706.5 | 4881.5 | 38956.0 | 47474.0 |
| 2012 | 932.0 | 2752.0 | 4799.0 | 39379.0 | 47862.0 |
| 2013 | 919.5 | 2742.0 | 4749.0 | 39862.0 | 48272.5 |
| 2014 | 903.5 | 2743.5 | 4656.5 | 40535.5 | 48839.0 |
| 2015 | 901.0 | 2727.0 | 4561.5 | 41179.0 | 49368.5 |

Table 12 shows the incidence rates of doctor-diagnosed HF between 2000-2015 in males by age at diagnosis, and Table 13 shows the corresponding data for females. Incidence rates have halved and we have also shown a decrease in HF admission rates over this period.[65] Some of this decrease may have been driven by decreases in CHD incidence. Most published national studies of trends in CHD incidence are from Scandinavian registry data. Mannsverk et al showed a steady downwards trend in CHD incidence in Sweden over time in both males and females from 1994 to 2010.[66] The age-and sex-adjusted incidence of total coronary heart disease decreased by 3% (95% CI 2.0% to 4.0%, P<0.001) each year. This decline was driven by decreases in out-hospital sudden death and hospitalized ST-elevation myocardial infarction. In Denmark, age-standardised incidence rates declined significantly from 2000 to 2009 for both sexes (females 445 to 340/100,000, males 822 to 678/100,000), reflecting a reduction in the annual number of new cases from 19.345 to 16.757.[67]

Table : incidence rates of doctor-diagnosed HF 2000-2015 in males by age at diagnosis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 16-44 | 45-64 | 65-74 | 75+ | All ages |
| 2000 | 71.5 | 940.5 | 1447.0 | 2384 | 4843.0 |
| 2001 | 84.5 | 954.0 | 1398.0 | 2456.5 | 4893.0 |
| 2002 | 107 | 954.5 | 1426.5 | 2390 | 4878.0 |
| 2003 | 117.0 | 982.5 | 1252 | 2277.0 | 4628.5 |
| 2004 | 127.5 | 991.5 | 1335.5 | 2037.5 | 4492.0 |
| 2005 | 125.5 | 912.5 | 1057.5 | 1906.0 | 4001.5 |
| 2006 | 108.5 | 848.5 | 1005 | 1631.5 | 3593.5 |
| 2007 | 119.0 | 824.5 | 944.5 | 1552.0 | 3440.0 |
| 2008 | 110.0 | 825.0 | 949 | 1631.0 | 3515.0 |
| 2009 | 107.5 | 783.5 | 883 | 1754.0 | 3528.0 |
| 2010 | 116.5 | 831.0 | 950.5 | 1651.5 | 3549.5 |
| 2011 | 105.5 | 790.5 | 906.5 | 1659.5 | 3462.0 |
| 2012 | 103.5 | 822.5 | 868.5 | 1708.5 | 3503.0 |
| 2013 | 100.5 | 734.0 | 850.5 | 1689.0 | 3374.0 |
| 2014 | 79.0 | 623.0 | 771.5 | 1512.5 | 2986.0 |
| 2015 | 75.5 | 564.5 | 650.5 | 1278.5 | 2569.0 |

Table . incidence rates of doctor-diagnosed HF 2000-2015 in females by age at diagnosis

|  | 16-44 | 45-64 | 65-74 | 75+ | All ages |
| --- | --- | --- | --- | --- | --- |
| 2000 | 32.0 | 439.0 | 1117.0 | 3758.0 | 5346.0 |
| 2001 | 57.5 | 457.5 | 1083.5 | 3687.5 | 5286.0 |
| 2002 | 45.0 | 419.5 | 1037.0 | 3585.5 | 5087.0 |
| 2003 | 45.0 | 419.5 | 898.5 | 3327.0 | 4690.0 |
| 2004 | 78.0 | 400.0 | 786.5 | 2722.5 | 3987.0 |
| 2005 | 39.0 | 313.0 | 697.0 | 2518.5 | 3567.5 |
| 2006 | 56.5 | 359.0 | 605.5 | 2118.0 | 3139.0 |
| 2007 | 57.5 | 293.5 | 599.5 | 1967.5 | 2918.0 |
| 2008 | 51.5 | 313.0 | 589.5 | 1925.0 | 2879.0 |
| 2009 | 73.0 | 322.5 | 550.5 | 2019.0 | 2965.0 |
| 2010 | 66.0 | 320.5 | 549.5 | 1933.0 | 2869.0 |
| 2011 | 59.0 | 325.0 | 524.5 | 1932.5 | 2841.0 |
| 2012 | 59.0 | 349.0 | 535.0 | 1881.0 | 2824.0 |
| 2013 | 45.0 | 290.0 | 503.5 | 1714.5 | 2553.0 |
| 2014 | 42.5 | 279.5 | 464.5 | 1645.5 | 2432.0 |
| 2015 | 53.5 | 239.5 | 393.0 | 1385.0 | 2071.0 |

### Logistic regression modelling

To reduce data volumes, the HF data we used for regression modelling is a case-control design, which required us to correct for sampling fraction. Missing ethnicity values were replaced by White.[18] Risk factors with low or minimal missingness were age, gender, ethnicity (see assumption above), CHD, diabetes, atrial fibrillation, and hypertension. The sometimes or often missing risk factors are BMI, smoking, and alcohol consumption. For these predictor variables the missing values were replaced by the inverse probability weight method.

The logistic regression model was fitted by using the product weights generated from the sampling probability weights multipliers, using the inverse probability weights method. Table 14 shows the multiple logistic regression model for the combined GP +HES diagnosis HF outcome, which was then utilised for the local prevalence estimates.

Table : multiple logistic regression model for the combined HF outcome GP+HES HF outcome plus probable cases

| Risk factors | ORs | P value | 95% CIs |
| --- | --- | --- | --- |
| Gender#Agegroup |  |  |  |
| Male#16-44 | 0.0007871 | <0.001 | [0.0007 - 0.0008] |
| Male#45-64 | 0.0071812 | <0.001 | [0.0069 - 0.0074] |
| Male#65-74 | 0.0227965 | <0.001 | [0.0221 - 0.0235] |
| Male#75+ | 0.0498207 | <0.001 | [0.0485 - 0.0512] |
| Female#16-44 | 0.0005963 | <0.001 | [0.0005 - 0.0006] |
| Female#45-64 | 0.0050637 | <0.001 | [0.0049 - 0.0052] |
| Female#65-74 | 0.0189998 | <0.001 | [0.0184 - 0.0196] |
| Female#75+ | 0.0499102 | <0.001 | [0.0487 - 0.0511] |
| BMI |  |  |  |
| <18.5 | 1.313427 | <0.001 | [1.2578 - 1.3715] |
| (18.5, 25] | 1.00 |  |  |
| (25, 30] | 1.061192 | <0.001 | [1.0422 - 1.0804] |
| >30 | 1.63105 | <0.001 | [1.5982 - 1.6645] |
| Smoking |  |  |  |
| Non-smoker | 1.00 |  |  |
| Smoker | 0.9712379 | 0.006 | [0.9511 - 0.9917] |
| Ex-smoker | 1.409611 | <0.001 | [1.3853 - 1.4342] |
| Alcohol consumption |  |  |  |
| Non-drinker | 1.00 |  |  |
| (0, 14] units/week | 0.922642 | <0.001 | [0.9040 - 0.9416] |
| (14, 42] units/week | 0.8956903 | <0.001 | [0.8743 - 0.9175] |
| >42 units/week | 1.069678 | 0.006 | [1.0193 - 1.1225] |
| Ethnicity |  |  |  |
| White | 1.00 |  |  |
| Black | 1.205105 | <0.001 | [1.0940 - 1.3274] |
| Asian | 1.220812 | <0.001 | [1.1260 - 1.3235] |
| Mixed | 1.185719 | <0.001 | [1.1572 - 1.2148] |
| Other | 1.037549 | 0.634 | [0.8915 - 1.2074] |
| CHD |  |  |  |
| 0 | 1.00 |  |  |
| Case | 2.658932 | <0.001 | [2.5996 - 2.7196] |
| Hypertension |  |  |  |
| 0 | 1.00 |  |  |
| Case | 1.447945 | <0.001 | [1.4242 - 1.4720] |
| Diabetes |  |  |  |
| 0 | 1.00 |  |  |
| Case  | 1.414926 | <0.001 | [1.3883 - 1.4420] |
| Atrial Fibrillation |  |  |  |
| 0 | 1.00 |  |  |
| Case | 5.220019 | <0.001 | [5.0962 - 5.3467] |

## External validation – Bland Altman plots

We externally validated the model-estimated prevalence by carrying out a disagreement analysis between model-estimated and QOF prevalence (%) of HF in practices. We estimated three principal components of disagreement (discordance as measured by Kendall's tau-a, bias as measured by median difference, and calibration as measured by the Theil-Sen median slope). Kendall's tau-a between mean prevalence and prevalence difference of heart failure at practice level is 0.254 (95% CIs (0.239 – 0.270, p<0.001), showing that prevalence means and model-QOF differences are 25.4 percent more likely to be concordant than to be discordant. Table 15 shows the percentile differences between practice-level model-estimated and QOF prevalence of heart failure.

Table : Percentile differences between model-estimated and QOF prevalence of heart failure

|  |  |  |
| --- | --- | --- |
| Percent | Percentile | (95% CI) |
| 0 | -2.3 | (-2.3, -2.3) |
| 25 | 0.4 | (0.4, 0.4) |
| 50 | 0.6 | (0.6, 0.6) |
| 75 | 0.9 | (0.9, 0.9) |
| 100 | 15.9 | (5.9, 5.9) |

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, using Bland-Altman plots. Figure 3 shows the Bland-Altman plot for the practice-level QOF and estimated prevalence for HF with higher modelled prevalence compared with QOF registered prevalence. Figure 4 is a scatter plot of practice-level model-estimated and QOF prevalence of diagnosed HF. In both cases, GP-diagnosed prevalence is lower than CPRD prevalence as defined above.

Figure : Bland-Altman plot for practice-level model-estimated and QOF prevalence of heart failure



Figure : Scatter plot of practice -level model-estimated and QOF prevalence of diagnosed heart failure



# References

1. Davies M, Hobbs F, Davis R, Kenkre J, Roalfe AK, Hare R, Wosornu D, Lancashire RJ. Prevalence of left-ventricular systolic dysfunction and heart failure in the Echocardiographic Heart of England Screening study: a population based study. Lancet 2001;358(9280):439-44. Link: <http://www.sciencedirect.com/science/article/pii/S0140673601056203>

<http://ac.els-cdn.com/S0140673601056203/1-s2.0-S0140673601056203-main.pdf?_tid=9516ec9e-f753-11e5-943a-00000aab0f01&acdnat=1459437586_c7dfb5563afe452f37eb11846ba31fd6>.

2. Bhatnagar P, Wickramasinghe K, Williams J, Rayner M, Townsend N. The epidemiology of cardiovascular disease in the UK 2014. Heart 2015;101(15):1182-9. Link: <http://heart.bmj.com/content/101/15/1182.full.pdf>.

3. Titler MG, Jensen GA, Dochterman JM, Xie XJ, Kanak M, Reed D, Shever LL. Cost of hospital care for older adults with heart failure: medical, pharmaceutical, and nursing costs. Health services research 2008;43(2):635-55. Link: <http://onlinelibrary.wiley.com/doi/10.1111/j.1475-6773.2007.00789.x/full>.

4. Voigt J, Sasha John M, Taylor A, Krucoff M, Reynolds MR, Michael Gibson C. A Reevaluation of the Costs of Heart Failure and Its Implications for Allocation of Health Resources in the United States. Clinical Cardiology 2014:n/a-n/a. Link: <http://dx.doi.org/10.1002/clc.22260>.

5. Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJV, Ponikowski P, Poole-Wilson PA, Strömberg A, van Veldhuisen DJ, Atar D, Hoes AW, Keren A, Mebazaa A, Nieminen M, Priori SG, Swedberg K. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart. European Heart Journal 2008;29(19):2388-442. Link: <http://eurheartj.oxfordjournals.org/content/ehj/29/19/2388.full.pdf>.

6. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failureThe Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) European Heart Journal 2016;37(27):2129-200. Link: <http://dx.doi.org/10.1093/eurheartj/ehw128>.

7. Braunschweig F, Cowie MR, Auricchio A. What are the costs of heart failure? Europace 2011;13(suppl 2):ii13-ii17. Link: <http://europace.oxfordjournals.org/content/13/suppl_2/ii13.abstract>.

8. Cook C, Cole G, Asaris P, Jabbour R, Francis D. 49 The Annual Global Economic Burden of Heart Failure. Heart 2014;100(Suppl 3):A28-A29. Link: <http://heart.bmj.com/content/100/Suppl_3/A28.2.abstract>.

9. Roger VL, Weston Sa, Redfield MM, Hellermann-Homan JP, Killian J, Yawn BP, Jacobsen SJ. Trends in heart failure incidence and survival in a community-based population. JAMA : the journal of the American Medical Association 2004;292(3):344-50. Link: <http://jama.jamanetwork.com/data/Journals/JAMA/4936/JOC32539.pdf>.

10. Taylor CJ, Ryan R, Nichols L, Gale N, Hobbs R, Marshall T. Survival following a diagnosis of heart failure in primary care. Family Practice 2017:8. Link: <https://academic.oup.com/fampra/article/2958553/Survival-following-a-diagnosis-of-heart-failure-in>.

11. Avery CL, Loehr LR, Baggett C, Chang PP, Kucharska-Newton AM, Matsushita K, Rosamond WD, Heiss G. The population burden of heart failure attributable to modifiable risk factors: the ARIC (Atherosclerosis Risk in Communities) study. J Am Coll Cardiol 2012;60(17):1640-6. Link: <http://ac.els-cdn.com/S0735109712028999/1-s2.0-S0735109712028999-main.pdf?_tid=5ea08e4a-ea9c-11e5-93df-00000aacb362&acdnat=1458039482_45d8a0e1fac958322d3c58e816646a69>.

12. Mitchell P, Marle D, Donkor A, Shote A, McDonagh T, et al. National Heart Failure Audit Report 2013-14. 2015. Link: <https://www.ucl.ac.uk/nicor/audits/heartfailure/documents/annualreports/hfannual13-14-updated.pdf>.

13. Gottdiener JS, Arnold AM, Aurigemma GP, Polak JF, Tracy RP, Kitzman DW, Gardin JM, Rutledge JE, Boineau RC. Predictors of congestive heart failure in the elderly: the Cardiovascular Health Study. J Am Coll Cardiol 2000;35(6):1628-37. Link: <http://ac.els-cdn.com/S0735109700005829/1-s2.0-S0735109700005829-main.pdf?_tid=f2b1d4b6-f809-11e5-8625-00000aacb361&acdnat=1459515913_217b3753f0c3f2e6f4874a85267144b1>.

14. Kenchaiah S, Narula J, Vasan RS. Risk factors for heart failure. The Medical clinics of North America 2004;88(5):1145-72. Link: <http://ac.els-cdn.com/S0025712504000665/1-s2.0-S0025712504000665-main.pdf?_tid=697e8f92-ea9c-11e5-9445-00000aab0f27&acdnat=1458039501_6d6371e3ef25d8f3ed1db7073d7b4dd2>.

15. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nat Rev Cardiol 2016. Link: <http://www.ncbi.nlm.nih.gov/pubmed/26935038>

<http://www.nature.com/nrcardio/journal/vaop/ncurrent/pdf/nrcardio.2016.25.pdf>.

16. Del Gobbo LC, Kalantarian S, Imamura F, Lemaitre R, Siscovick DS, Psaty BM, Mozaffarian D. Contribution of Major Lifestyle Risk Factors for Incident Heart Failure in Older Adults: The Cardiovascular Health Study. JACC. Heart failure 2015;3(7):520-8. Link: <http://ac.els-cdn.com/S2213177915002140/1-s2.0-S2213177915002140-main.pdf?_tid=601751f6-f815-11e5-8459-00000aacb35f&acdnat=1459520819_f07a813e318ba4f88531e60186d534b8>.

17. Kannan A, Janardhanan R. Hypertension as a risk factor for heart failure. Current hypertension reports 2014;16(7):447. Link: [http://download.springer.com/static/pdf/621/art%253A10.1007%252Fs11906-014-0447-7.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs11906-014-0447-7&token2=exp=1458564688~acl=%2Fstatic%2Fpdf%2F621%2Fart%25253A10.1007%25252Fs11906-014-0447-7.pdf%3ForiginUrl%3Dhttp%253A%252F%252Flink.springer.com%252Farticle%252F10.1007%252Fs11906-014-0447-7\*~hmac=6effea38b9165c1a0b37a0f20160b51655981428b783a5c3ed5a69836bd2191d](http://download.springer.com/static/pdf/621/art%253A10.1007%252Fs11906-014-0447-7.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs11906-014-0447-7&token2=exp=1458564688~acl=%2Fstatic%2Fpdf%2F621%2Fart%25253A10.1007%25252Fs11906-014-0447-7.pdf%3ForiginUrl%3Dhttp%253A%252F%252Flink.springer.com%252Farticle%252F10.1007%252Fs11906-014-0447-7*~hmac=6effea38b9165c1a0b37a0f20160b51655981428b783a5c3ed5a69836bd2191d).

18. Hippisley-Cox J, Coupland C. Development and validation of risk prediction equations to estimate future risk of heart failure in patients with diabetes: a prospective cohort study. BMJ Open 2015;5(9):e008503. Link: <http://bmjopen.bmj.com/content/5/9/e008503>.

19. Clerfond G, Biere L, Mateus V, Grall S, Willoteaux S, Prunier F, Furber A. End-systolic wall stress predicts post-discharge heart failure after acute myocardial infarction. Archives of cardiovascular diseases 2015;108(5):310-20. Link: <http://ac.els-cdn.com/S1875213615000455/1-s2.0-S1875213615000455-main.pdf?_tid=8151ecd8-f80a-11e5-8767-00000aab0f6c&acdnat=1459516150_9a2b127974efd72ca236183c329c2d85>.

20. Hellermann JP, Jacobsen SJ, Redfield MM, Reeder GS, Weston SA, Roger VL. Heart failure after myocardial infarction: clinical presentation and survival. Eur J Heart Fail 2005;7(1):119-25. Link: <http://onlinelibrary.wiley.com/store/10.1016/j.ejheart.2004.04.011/asset/ejhf2004-04-011.pdf?v=1&t=im22p4oy&s=972e6cb9e6511663522ad692eb16984ae5ab9cae>.

21. Gjesing A, Gislason GH, Kober L, Gustav Smith J, Christensen SB, Gustafsson F, Olsen AM, Torp-Pedersen C, Andersson C. Nationwide trends in development of heart failure and mortality after first-time myocardial infarction 1997-2010: A Danish cohort study. European journal of internal medicine 2014;25(8):731-8. Link: <http://ac.els-cdn.com/S0953620514002441/1-s2.0-S0953620514002441-main.pdf?_tid=1990145c-f80b-11e5-b8e5-00000aab0f6c&acdnat=1459516406_e67f68d8bff50a7bd8d833bd61550c07>.

22. Ezekowitz JA, Kaul P, Bakal JA, Armstrong PW, Welsh RC, McAlister FA. Declining in-hospital mortality and increasing heart failure incidence in elderly patients with first myocardial infarction. J Am Coll Cardiol 2009;53(1):13-20. Link: <http://ac.els-cdn.com/S0735109708033159/1-s2.0-S0735109708033159-main.pdf?_tid=2b8d59e4-f80b-11e5-8448-00000aab0f02&acdnat=1459516436_f6ebb39d395168ce650781161d7bcea8>.

23. Gheorghiade M, Sopko G, De Luca L, Velazquez EJ, Parker JD, Binkley PF, Sadowski Z, Golba KS, Prior DL, Rouleau JL, Bonow RO. Navigating the crossroads of coronary artery disease and heart failure. Circulation 2006;114(11):1202-13. Link: <http://circ.ahajournals.org/content/114/11/1202.full.pdf>.

24. Liu C, Hou Y, Wang X, Zhao Z, Liu Z, Zhai J, Mao J, Shang H. Clinical assessment of Shenfu injection loading in the treatment of patients with exacerbation of chronic heart failure due to coronary heart disease: study protocol for a randomized controlled trial. Trials 2015;16:222. Link: <http://trialsjournal.biomedcentral.com/articles/10.1186/s13063-015-0729-7>.

25. Rodriguez FH, 3rd, Moodie DS, Parekh DR, Franklin WJ, Morales DL, Zafar F, Adams GJ, Friedman RA, Rossano JW. Outcomes of heart failure-related hospitalization in adults with congenital heart disease in the United States. Congenital heart disease 2013;8(6):513-9. Link: <http://onlinelibrary.wiley.com/store/10.1111/chd.12019/asset/chd12019.pdf?v=1&t=imhqbu2m&s=ad0dcf3e8e334ecb14abdb2b981b9afd10a54fd3>.

26. Corrao G, Rea F, Ghirardi A, Soranna D, Merlino L, Mancia G. Adherence with antihypertensive drug therapy and the risk of heart failure in clinical practice. Hypertension 2015;66(4):742-9. Link: <http://hyper.ahajournals.org/content/66/4/742.full.pdf>.

27. Lala A, Desai AS. The role of coronary artery disease in heart failure. Heart failure clinics 2014;10(2):353-65. Link: <http://ac.els-cdn.com/S155171361300113X/1-s2.0-S155171361300113X-main.pdf?_tid=b2ce213c-ef6d-11e5-853d-00000aacb360&acdnat=1458569193_726a601e0022611b69b49c74fd97542f>.

28. Silverman MG, Patel B, Blankstein R, Lima JA, Blumenthal RS, Nasir K, Blaha MJ. Impact of Race, Ethnicity, and Multimodality Biomarkers on the Incidence of New-Onset Heart Failure With Preserved Ejection Fraction (from the Multi-Ethnic Study of Atherosclerosis). Am J Cardiol 2016. Link: <http://ac.els-cdn.com/S0002914916302302/1-s2.0-S0002914916302302-main.pdf?_tid=8d9e794c-f80b-11e5-8f2d-00000aab0f26&acdnat=1459516601_dffb5a89601becbd43a82ef8aac2a382>.

29. Britton KA, Gaziano JM, Djousse L. Normal systolic blood pressure and risk of heart failure in US male physicians. Eur J Heart Fail 2009;11(12):1129-34. Link: <http://onlinelibrary.wiley.com/doi/10.1093/eurjhf/hfp141/epdf>.

30. Zanchetti A, Thomopoulos C, Parati G. Randomized controlled trials of blood pressure lowering in hypertension: a critical reappraisal. Circulation research 2015;116(6):1058-73. Link.

31. Turnbull F, Neal B, Pfeffer M, Kostis J, Algert C, Woodward M, Chalmers J, Zanchetti A, MacMahon S. Blood pressure-dependent and independent effects of agents that inhibit the renin-angiotensin system. Journal of hypertension 2007;25(5):951-8. Link.

32. Chirinos JA, Segers P, Duprez DA, Brumback L, Bluemke DA, Zamani P, Kronmal R, Vaidya D, Ouyang P, Townsend RR, Jacobs DR, Jr. Late systolic central hypertension as a predictor of incident heart failure: the Multi-ethnic Study of Atherosclerosis. Journal of the American Heart Association 2015;4(3):e001335. Link: <http://jaha.ahajournals.org/content/4/3/e001335.long>.

33. Rosengren A, Vestberg D, Svensson AM, Kosiborod M, Clements M, Rawshani A, Pivodic A, Gudbjornsdottir S, Lind M. Long-term excess risk of heart failure in people with type 1 diabetes: a prospective case-control study. The lancet. Diabetes & endocrinology 2015;3(11):876-85. Link: <http://ac.els-cdn.com/S2213858715002922/1-s2.0-S2213858715002922-main.pdf?_tid=79030dfe-ef82-11e5-93c4-00000aab0f27&acdnat=1458578116_d1e434e9990fb79f6774880f38e83f06>.

34. Wang Y, Negishi T, Negishi K, Marwick TH. Prediction of heart failure in patients with type 2 diabetes mellitus- a systematic review and meta-analysis. Diabetes research and clinical practice 2015;108(1):55-66. Link: <http://ac.els-cdn.com/S0168822715000194/1-s2.0-S0168822715000194-main.pdf?_tid=3afc2a04-f815-11e5-8459-00000aacb35f&acdnat=1459520757_0f0f19e7c4e8d4f2d9eeab3d483e4d1f>.

35. Schmidt M, Botker HE, Pedersen L, Sorensen HT. Young adulthood obesity and risk of acute coronary syndromes, stable angina pectoris, and congestive heart failure: a 36-year cohort study. Annals of epidemiology 2014;24(5):356-61.e1. Link: <http://ac.els-cdn.com/S1047279714000416/1-s2.0-S1047279714000416-main.pdf?_tid=75865528-f815-11e5-8635-00000aacb35f&acdnat=1459520855_3419e74703b568ff36896e4818d6fed5>.

36. Gopal DM, Kalogeropoulos AP, Georgiopoulou VV, Smith AL, Bauer DC, Newman AB, Kim L, Bibbins-Domingo K, Tindle H, Harris TB, Tang WW, Kritchevsky SB, Butler J. Cigarette smoking exposure and heart failure risk in older adults: the Health, Aging, and Body Composition Study. Am Heart J 2012;164(2):236-42. Link: <http://ac.els-cdn.com/S0002870312003699/1-s2.0-S0002870312003699-main.pdf?_tid=664c91da-ef78-11e5-89b5-00000aab0f01&acdnat=1458573789_fec2e5acc17690e70432a4ac049e8db2>.

37. Ahmed AA, Patel K, Nyaku MA, Kheirbek RE, Bittner V, Fonarow GC, Filippatos GS, Morgan CJ, Aban IB, Mujib M, Desai RV, Allman RM, White M, Deedwania P, Howard G, Bonow RO, Fletcher RD, Aronow WS, Ahmed A. Risk of Heart Failure and Death After Prolonged Smoking Cessation: Role of Amount and Duration of Prior Smoking. Circ Heart Fail 2015;8(4):694-701. Link: <http://circheartfailure.ahajournals.org/content/8/4/694.full.pdf>.

38. Wannamethee SG, Shaper AG, Papacosta O, Lennon L, Welsh P, Whincup PH. Lung function and airway obstruction: associations with circulating markers of cardiac function and incident heart failure in older men-the British Regional Heart Study. Thorax 2016. Link: <http://thorax.bmj.com/content/early/2016/01/25/thoraxjnl-2014-206724.full.pdf>.

39. Larsson SC, Orsini N, Wolk A. Alcohol consumption and risk of heart failure: a dose-response meta-analysis of prospective studies. Eur J Heart Fail 2015;17(4):367-73. Link: <http://onlinelibrary.wiley.com/store/10.1002/ejhf.228/asset/ejhf228.pdf?v=1&t=imht5h7x&s=a7d5593d75a10721773c48250ce5b9f7187e4102>.

40. Ljunggren M, Byberg L, Theorell-Haglow J, Lindahl B, Michaelsson K, Lindberg E. Increased risk of heart failure in women with symptoms of sleep-disordered breathing. Sleep medicine 2016;17:32-7. Link: <http://www.ncbi.nlm.nih.gov/pubmed/26847971>

<http://ac.els-cdn.com/S1389945715020092/1-s2.0-S1389945715020092-main.pdf?_tid=a1e9e636-ea9e-11e5-97fd-00000aacb35d&acdnat=1458040454_62ef419747446699fca39a725eda51b5>

<http://ac.els-cdn.com/S1389945715020092/1-s2.0-S1389945715020092-main.pdf?_tid=a48355d0-ea9e-11e5-807f-00000aacb360&acdnat=1458040459_690d5ec5adaf59a784a4924aed5c12a8>

<http://ac.els-cdn.com/S1389945715020092/1-s2.0-S1389945715020092-main.pdf?_tid=b090ecc0-ea9e-11e5-80fd-00000aab0f02&acdnat=1458040479_7552709dc26870fbfd5be32c405683b8>.

41. Turnbull F. Effects of different blood-pressure-lowering regimens on major cardiovascular events: results of prospectively-designed overviews of randomised trials. Lancet 2003;362(9395):1527-35. Link: <http://ac.els-cdn.com/S0140673603147393/1-s2.0-S0140673603147393-main.pdf?_tid=1af1f4d0-f817-11e5-87d3-00000aacb360&acdnat=1459521562_ab5df065dfbed25751999d7905fa6cf1>.

42. Heist EK, Ruskin JN. Atrial fibrillation and congestive heart failure: risk factors, mechanisms, and treatment. Progress in cardiovascular diseases 2006;48(4):256-69. Link: <http://ac.els-cdn.com/S0033062005001118/1-s2.0-S0033062005001118-main.pdf?_tid=c8a53072-fc01-11e5-aa78-00000aacb35e&acdnat=1459952209_592881d9bcfa04d2a3efd14a554b19b9>.

43. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. Heart 2007;93(9):1137-46. Link: <http://heart.bmj.com/content/93/9/1137.long>.

44. Velavan P, Khan NK, Goode K, Rigby AS, Loh PH, Komajda M, Follath F, Swedberg K, Madeira H, Cleland JG. Predictors of short term mortality in heart failure - insights from the Euro Heart Failure survey. International journal of cardiology 2010;138(1):63-9. Link: <http://ac.els-cdn.com/S0167527308009509/1-s2.0-S0167527308009509-main.pdf?_tid=e7b3d8a2-f818-11e5-ad5a-00000aacb35e&acdnat=1459522335_3ad8518d7bfd9ea9c8bf69a52bf967cf>.

45. Zomer AC, Vaartjes I, van der Velde ET, de Jong HM, Konings TC, Wagenaar LJ, Heesen WF, Eerens F, Baur LH, Grobbee DE, Mulder BJ. Heart failure admissions in adults with congenital heart disease; risk factors and prognosis. International journal of cardiology 2013;168(3):2487-93. Link: <http://ac.els-cdn.com/S016752731300421X/1-s2.0-S016752731300421X-main.pdf?_tid=81fad850-ea9c-11e5-a9be-00000aacb360&acdnat=1458039542_11bfa0c24b85060744f69904cdf4d66c>.

46. Verheugt CL, Uiterwaal CS, Grobbee DE, Mulder BJ. Long-term prognosis of congenital heart defects: a systematic review. International journal of cardiology 2008;131(1):25-32. Link: <http://ac.els-cdn.com/S0167527308007936/1-s2.0-S0167527308007936-main.pdf?_tid=70d58428-f819-11e5-9dda-00000aacb35e&acdnat=1459522565_8949aeb2542cb84ab39119da242c9f9d>.

47. Dolk H, Loane M, Garne E. Congenital heart defects in Europe: prevalence and perinatal mortality, 2000 to 2005. Circulation 2011;123(8):841-9. Link: <http://circ.ahajournals.org/content/123/8/841.full.pdf>.

48. Meyer S, Brouwers FP, Voors AA, Hillege HL, de Boer RA, Gansevoort RT, van der Harst P, Rienstra M, van Gelder IC, van Veldhuisen DJ, van Gilst WH, van der Meer P. Sex differences in new-onset heart failure. Clinical research in cardiology : official journal of the German Cardiac Society 2015;104(4):342-50. Link: [http://download.springer.com/static/pdf/652/art%253A10.1007%252Fs00392-014-0788-x.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs00392-014-0788-x&token2=exp=1459953526~acl=%2Fstatic%2Fpdf%2F652%2Fart%25253A10.1007%25252Fs00392-014-0788-x.pdf%3ForiginUrl%3Dhttp%253A%252F%252Flink.springer.com%252Farticle%252F10.1007%252Fs00392-014-0788-x\*~hmac=7c51474bc65e543cb549c92a3f642720843a7bd87465590d1f76bf26d3de93df](http://download.springer.com/static/pdf/652/art%253A10.1007%252Fs00392-014-0788-x.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs00392-014-0788-x&token2=exp=1459953526~acl=%2Fstatic%2Fpdf%2F652%2Fart%25253A10.1007%25252Fs00392-014-0788-x.pdf%3ForiginUrl%3Dhttp%253A%252F%252Flink.springer.com%252Farticle%252F10.1007%252Fs00392-014-0788-x*~hmac=7c51474bc65e543cb549c92a3f642720843a7bd87465590d1f76bf26d3de93df).

49. Wannamethee SG, Whincup PH, Lennon L, Papacosta O, Shaper AG. Alcohol consumption and risk of incident heart failure in older men: a prospective cohort study. Open heart 2015;2(1):e000266. Link: [http://openheart.bmj.com/content/2/1/e000266.full.pdf+html](http://openheart.bmj.com/content/2/1/e000266.full.pdf%2Bhtml).

50. Seferovic PM, Stoerk S, Filippatos G, Mareev V, Kavoliuniene A, Ristic AD, Ponikowski P, McMurray J, Maggioni A, Ruschitzka F, van Veldhuisen DJ, Coats A, Piepoli M, McDonagh T, Riley J, Hoes A, Pieske B, Dobric M, Papp Z, Mebazaa A, Parissis J, Ben Gal T, Vinereanu D, Brito D, Altenberger J, Gatzov P, Milinkovic I, Hradec J, Trochu JN, Amir O, Moura B, Lainscak M, Comin J, Wikstrom G, Anker S. Organization of heart failure management in European Society of Cardiology member countries: survey of the Heart Failure Association of the European Society of Cardiology in collaboration with the Heart Failure National Societies/Working Groups. Eur J Heart Fail 2013;15(9):947-59. Link: <http://onlinelibrary.wiley.com/store/10.1093/eurjhf/hft092/asset/ejhfhft092.pdf?v=1&t=imgmnhig&s=90b7f7fb3f30aa83ddbf89e4bf44c3d92b289551>.

51. Nick Townsend JW, Prachi Bhatnagar, Kremlin Wickramasinghe, Rayner. aM. Cardiovascular disease statistics. British Heart Foundation 2014. Link: <https://www.bhf.org.uk/~/media/files/publications/research/bhf_cvd-statistics-2014_web_2.pdf>.

52. National Clinical Guideline Centre, National Institute for Health and Care Excellence. Acute Heart Failure: Diagnosing and Managing Acute Heart Failure in Adults

London: 2014. <https://www.nice.org.uk/guidance/cg187?unlid=99534687420163361647>.

53. Swedberg K, Cleland J, Dargie H, Drexler H, Follath F, Komajda M, Tavazzi L, Smiseth Oa, Gavazzi A, Haverich A, Hoes A, Jaarsma T, Korewicki J, Lévy S, Linde C, Lopez-Sendon J-L, Nieminen MS, Piérard L, Remme WJ. Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005): The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology. European Heart Journal 2005;26(11):1115-40. Link: <http://www.ncbi.nlm.nih.gov/pubmed/15901669>.

54. Kotecha D, Manzano L, Krum H, Rosano G, Holmes J, Altman DG, Collins PD, Packer M, Wikstrand J, Coats AJS, Cleland JGF, Kirchhof P, von Lueder TG, Rigby AS, Andersson B, Lip GY, van Veldhuisen DJ, Shibata MC, Wedel H, Böhm M, Flather MD. Effect of age and sex on efficacy and tolerability of β blockers in patients with heart failure with reduced ejection fraction: individual patient data meta-analysis. BMJ 2016;353. Link: <http://www.bmj.com/content/bmj/353/bmj.i1855.full.pdf>.

55. Hoes AW. β blockers for heart failure. BMJ 2016;353. Link: <http://www.bmj.com/content/bmj/353/bmj.i2074.full.pdf>.

56. Nacul L, Soljak M, Samarasundera E, Hopkinson NS, Lacerda E, Indulkar T, Flowers J, Walford H, Majeed A. COPD in England: a comparison of expected, model-based prevalence and observed prevalence from general practice data. Journal of Public Health 2011;33(1):108-16. Link: <http://jpubhealth.oxfordjournals.org/content/33/1/108.full>.

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

58. Newson RB. Attributable and unattributable risks and fractions and other scenario comparisons. Stata Journal 2013;13(4):672-98. Link: <Go to ISI>://WOS:000329680600001.

59. Newson RB. Attributable and unattributable risks and fractions and other scenario comparisons. Stata J 2013;13:672-98. Link.

60. NHS Employers. Quality and Outcomes Framework guidance for GMS contract 2013/14. Online: Health, 2013. <http://www.nhsemployers.org/your-workforce/primary-care-contacts/general-medical-services/quality-and-outcomes-framework/changes-to-qof-201314>.

61. Soljak M, Samarasundera E, Indulkar T, Walford H, Majeed A. Variations in cardiovascular disease under-diagnosis in England: national cross-sectional spatial analysis. BMC Cardiovascular Disorders 2011;11(1):12. Link: <http://www.biomedcentral.com/1471-2261/11/12>.

62. Bland JM, Altman DG. Measuring agreement in method comparison studies. Statistical Methods in Medical Research 1999;8(2):135-60. Link: <http://smm.sagepub.com/content/8/2/135.abstract>.

63. Ludbrook J. Linear regression analysis for comparing two measurers or methods of measurement: But which regression? Clinical and Experimental Pharmacology and Physiology 2010;37(7):692-99. Link: <http://dx.doi.org/10.1111/j.1440-1681.2010.05376.x>.

64. Francq BG, Govaerts B. How to regress and predict in a Bland–Altman plot? Review and contribution based on tolerance intervals and correlated-errors-in-variables models. Statistics in Medicine 2016;35(14):2328-58. Link: <http://dx.doi.org/10.1002/sim.6872>.

65. Brettell R, Soljak M, Cecil E, Cowie MR, Tuppin P, Majeed A. Reducing heart failure admission rates in England 2004–2011 are not related to changes in primary care quality: national observational study. European Journal of Heart Failure 2013. Link: <http://eurjhf.oxfordjournals.org/content/early/2013/07/11/eurjhf.hft107.abstract>.

66. Mannsverk J, Wilsgaard T, Mathiesen EB, Løchen M-L, Rasmussen K, Thelle DS, Njølstad I, Hopstock LA, Bønaa KH. Trends in Modifiable Risk Factors are Associated With Declining Incidence of Hospitalized and Non-Hospitalized Acute Coronary Heart Disease in a Population. Circulation 2015. Link.

67. Koch MB, Davidsen M, Andersen LV, Juel K, Jensen GB. Increasing prevalence despite decreasing incidence of ischaemic heart disease and myocardial infarction. A national register based perspective in Denmark, 1980–2009. European Journal of Preventive Cardiology 2013;22(2):189-95. Link: <http://dx.doi.org/10.1177/2047487313509495>.