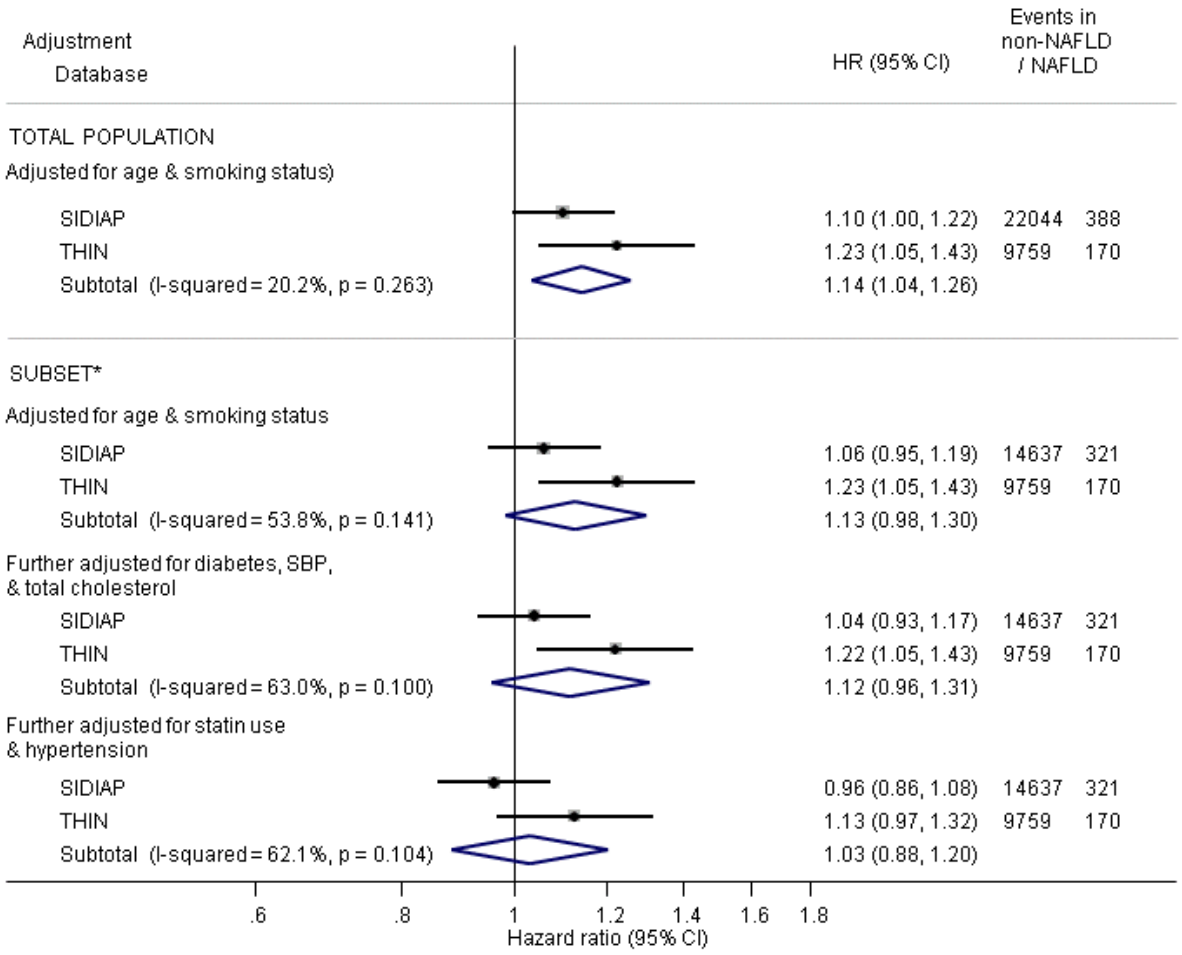
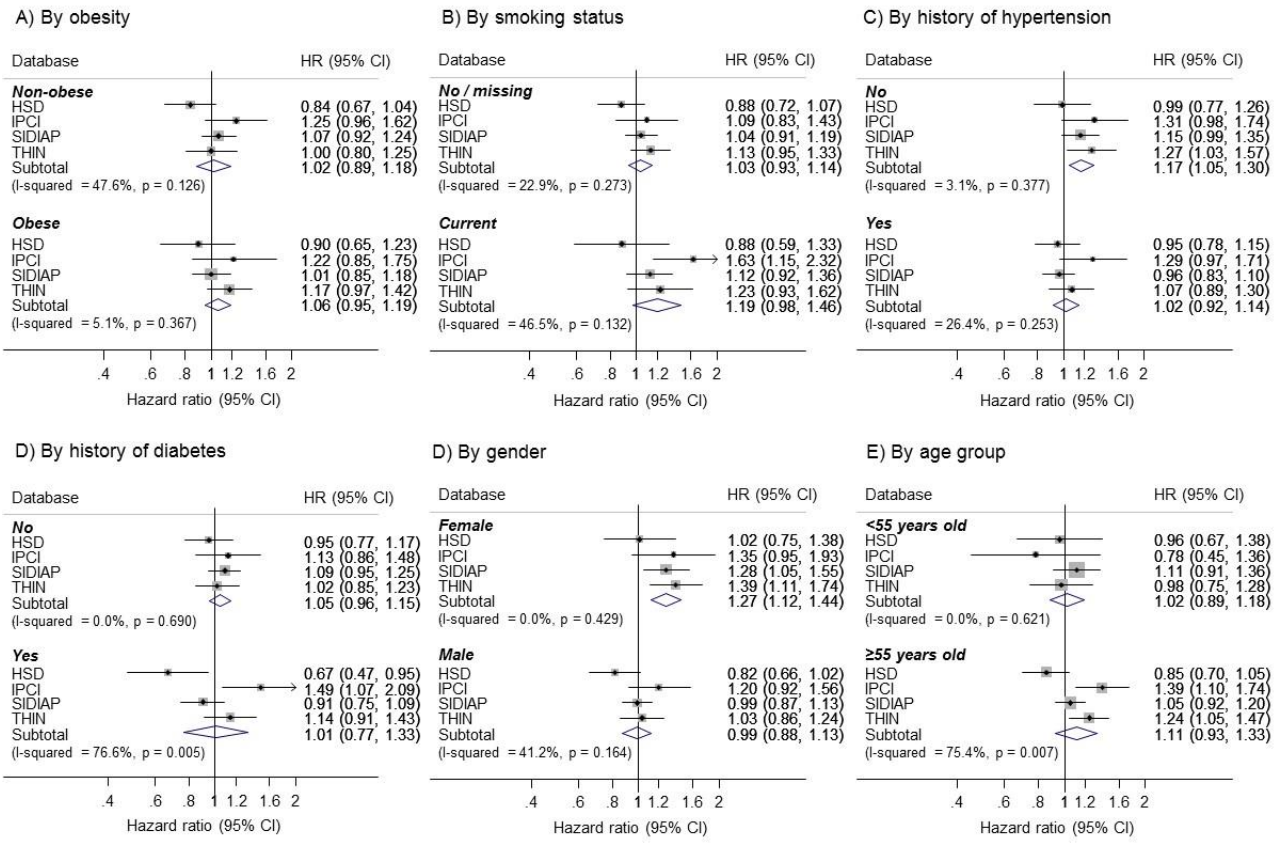
SUPPLEMENTARY MATERIAL



Supplementary Figure 1. Hazard ratios for myocardial infarction in NAFLD patients without a NASH records only. Note: Weights are from random-effect meta-analysis and inversely proportional to the variance of the estimated hazard ratios (therefore proportional to the number of events contributing the hazard ratios)..

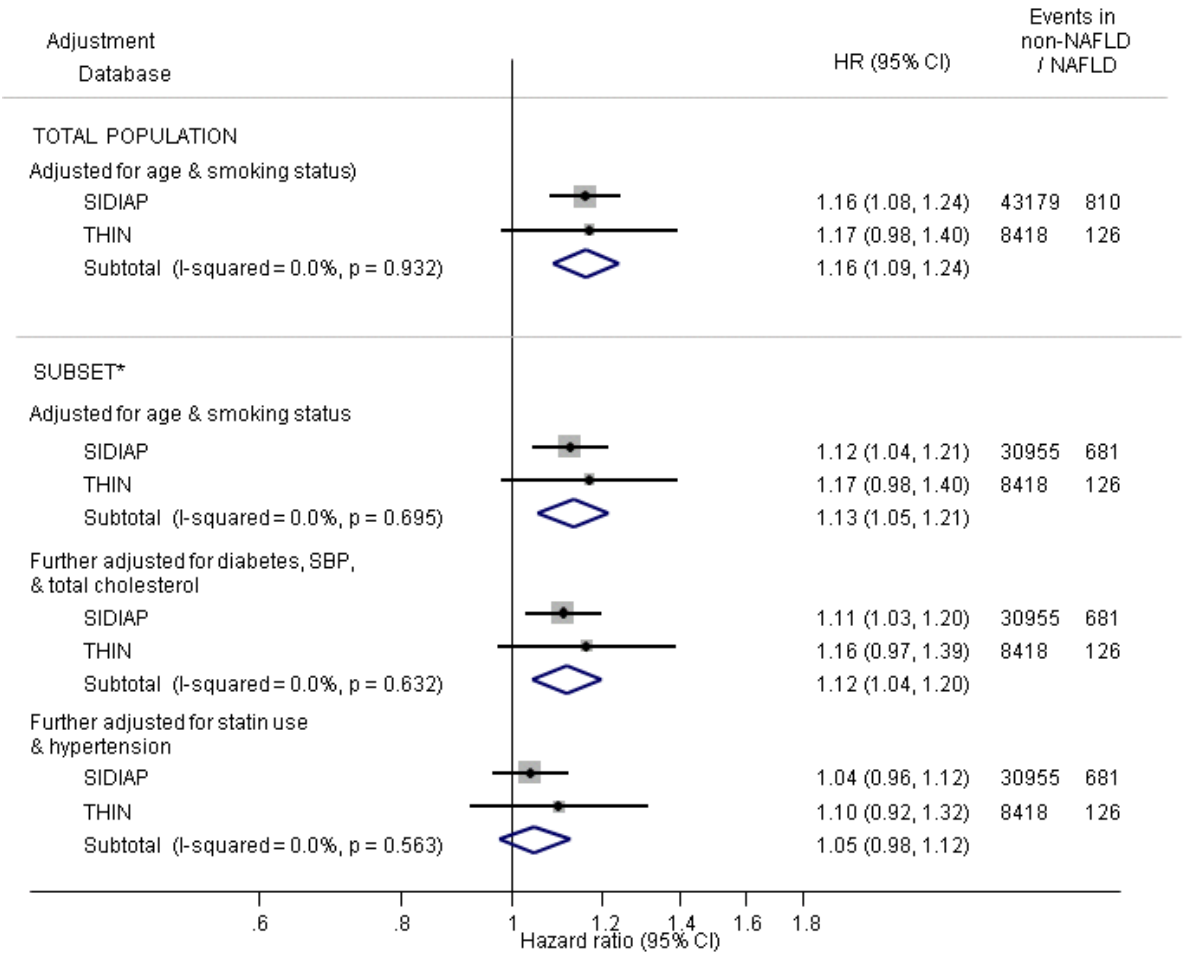
Data is presented as hazard ratio and their 95% CI. Analyses were progressively adjusted for age, smoking status, type 2 diabetes, SBP, total cholesterol, statin use and hypertension.

Note: it was only possible to run this analysis in the SIDIAP and THIN databases as separate codes are available for NAFLD and NASH.



Supplementary Figure 2. Hazard ratio for myocardial infarction in subgroup and pooled by multivariate meta-analysis. Note: Weights are from random-effect meta-analysis and inversely proportional to the variance of the estimated hazard ratios (therefore proportional to the number of events contributing the hazard ratios)..

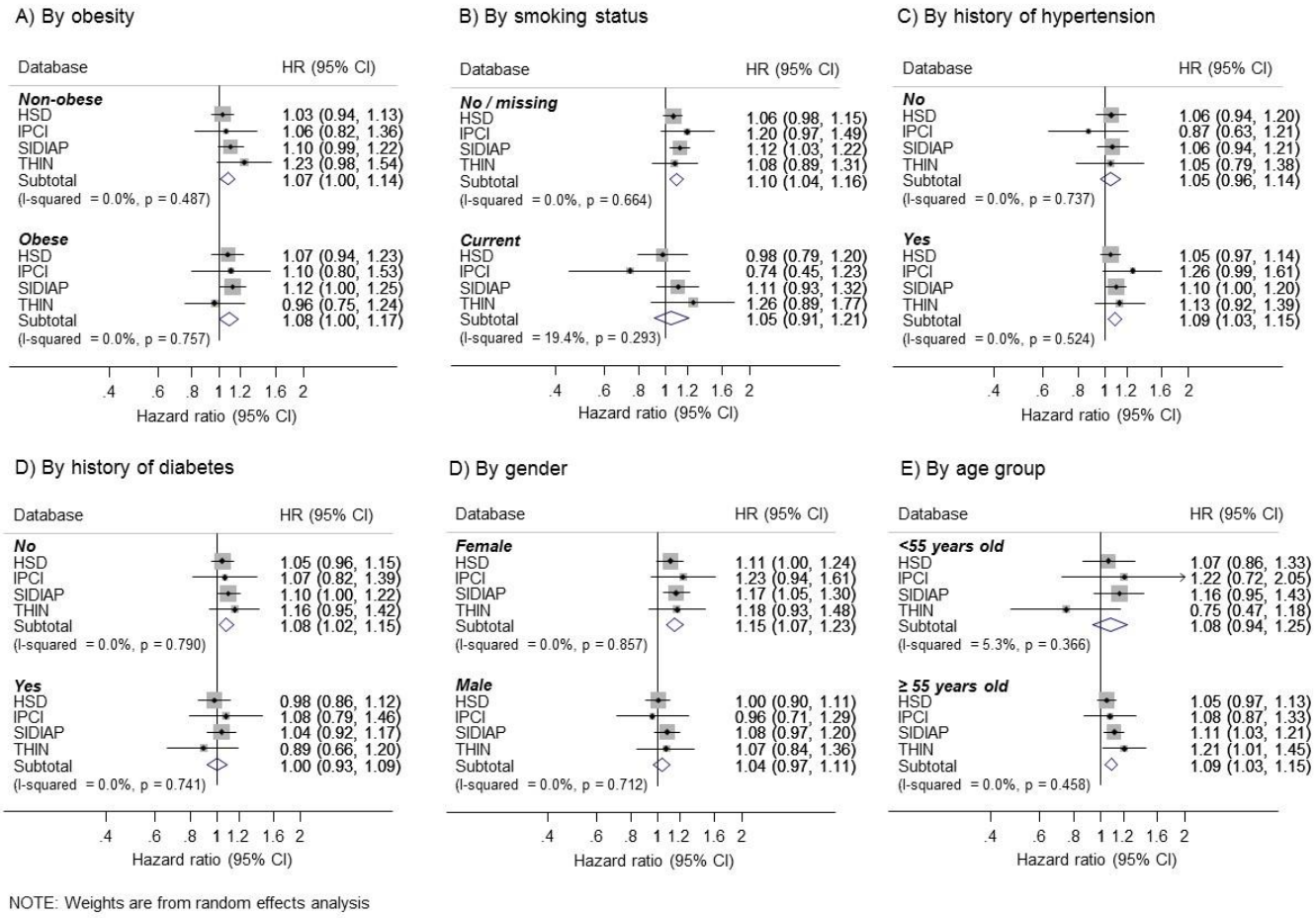
Data is presented as hazard ratio and their 95% CI. Analyses were adjusted for age, smoking status, SBP, total cholesterol. Estimates were pooled by random effects meta-analysis within each subgroup.



Supplementary Figure 3. Hazard ratio for stroke in NAFLD patients without a NASH records. Note: Weights are from random-effect meta-analysis and inversely proportional to the variance of the estimated hazard ratios (therefore proportional to the number of events contributing the hazard ratios)..

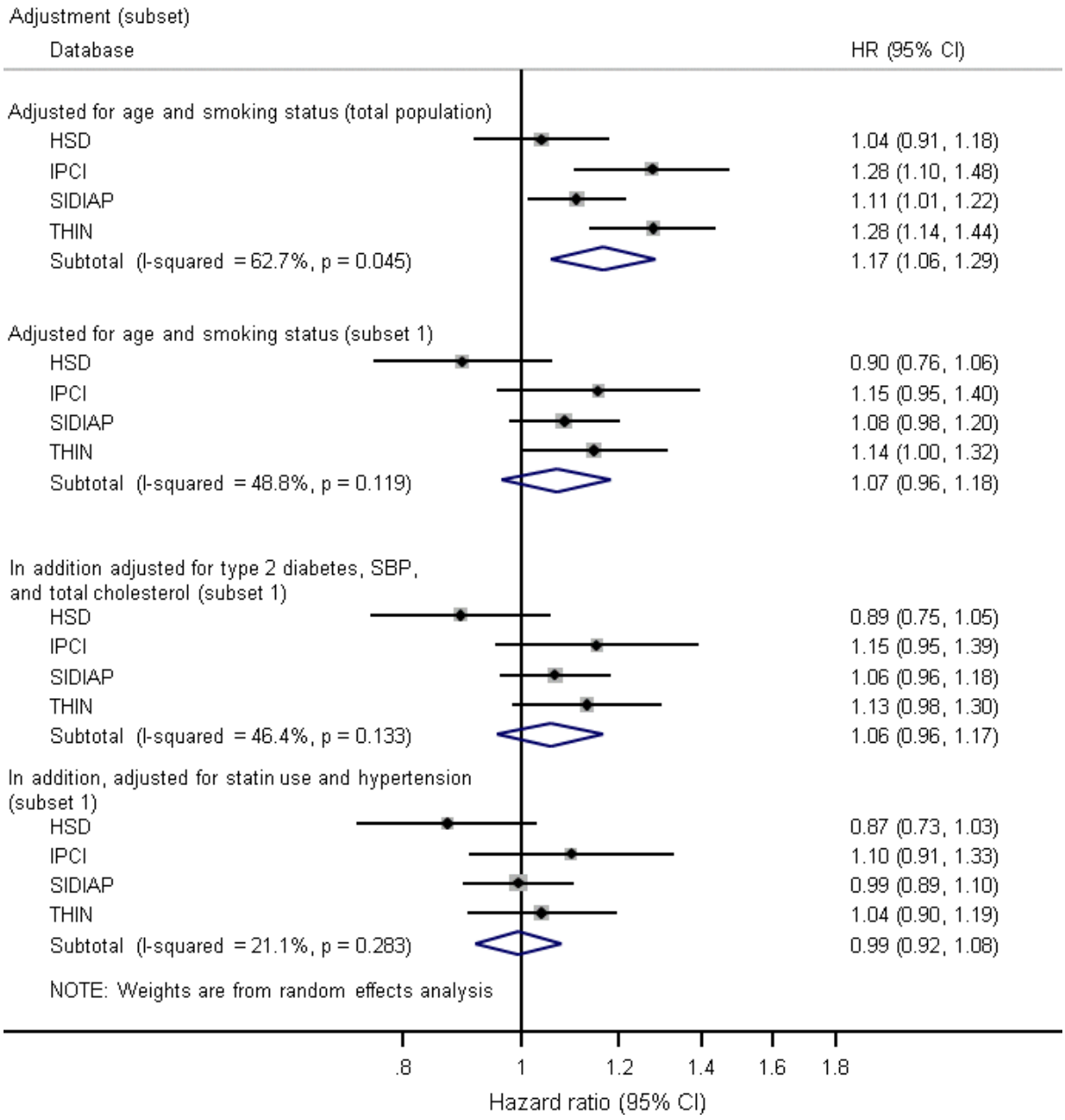
Data is presented as hazard ratio and their 95% CI. Analyses were progressively adjusted for age, smoking status, type 2 diabetes, SBP, total cholesterol, statin use and hypertension.

Note: it was only possible to run this analysis in the SIDIAP and THIN databases as separate codes are available for NAFLD and NASH.



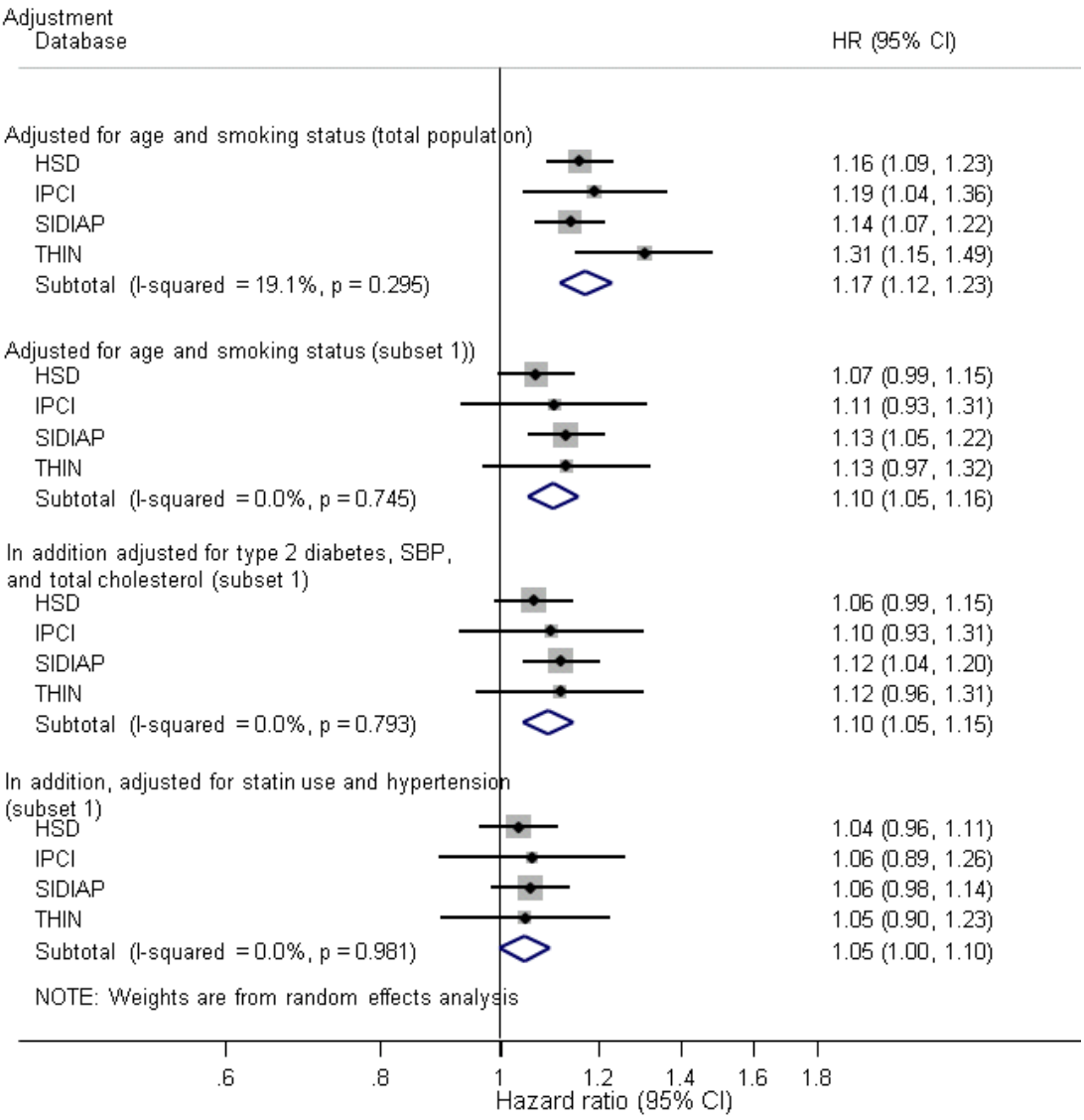
Supplementary Figure 4. Hazard ratio for stroke in NAFLD patients without a NASH records by subgroup and pooled across databases by multivariate meta-analysis. Note: Weights are from random-effect meta-analysis and inversely proportional to the variance of the estimated hazard ratios (therefore proportional to the number of events contributing the hazard ratios)..

Data is presented as hazard ratio and their 95% CI. Analyses adjusted for age, smoking status, SBP and total cholesterol. Estimates were pooled by random effects meta-analysis within each subgroup.



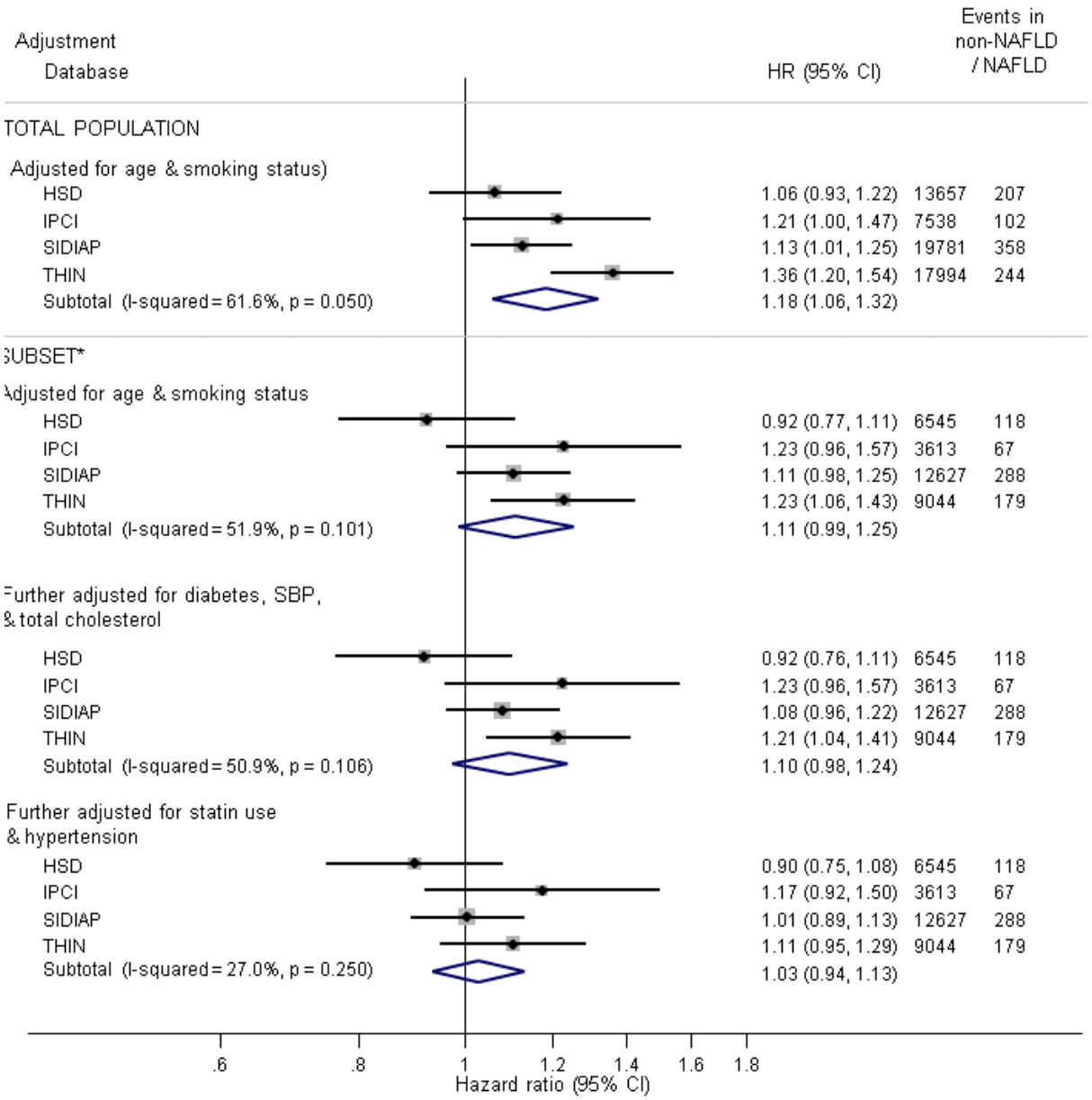
Supplementary Figure 5. Sensitivity analyses - Hazard ratio for myocardial infarction with in NAFLD patients including patients with less than 6 months of medical history prior and follow-up post index date, or who had a history of stroke or MI. Note: Weights are from random-effect meta-analysis and inversely proportional to the variance of the estimated hazard ratios (therefore proportional to the number of events contributing the hazard ratios)..

Data is presented as hazard ratio and their 95% CI. Analyses were progressively adjusted for age, smoking status, type 2 diabetes, SBP, total cholesterol, statin use and hypertension.



Supplementary Figure 6. Sensitivity analyses for hazard ratio for stroke in patients with NAFLD including patients with less than 6 months of medical history prior and follow-up post index date, or who had a history of stroke or MI. Note: Weights are from random-effect meta-analysis and inversely proportional to the variance of the estimated hazard ratios (therefore proportional to the number of events contributing the hazard ratios)..

Data is presented as hazard ratio and their 95% CI. Analyses were progressively adjusted for age, smoking status, type 2 diabetes, SBP, total cholesterol, statin use and hypertension.



Supplementary Figure 7. Association of NAFLD/NASH with myocardial infarction excluding patients with less than 6 months follow-up post index date (excluding events happening in the first 6 months after index date). Note: Weights are from random-effect meta-analysis and inversely proportional to the variance of the estimated hazard ratios (therefore proportional to the number of events contributing the hazard ratios).

Data is presented as hazard ratio and their 95% CI. Analyses were progressively adjusted for age, smoking status, type 2 diabetes, SBP, total cholesterol, statin use and hypertension. Data for age and smoking (total population data set) was available for 59,881 (patients without NAFLD n=58,970; patients with NAFLD n=911). A subset* of participants have full data available for age, smoking, type 2 diabetes, SBP, total cholesterol, statin use and hypertension, therefore the analyses were restricted to 32,481 (Non NAFLD patients n=31,829; patients with NAFLD n=652).

Supplementary Table 1 Attrition table

Attrition criteria	HSD (Italy)	IPCI (The Netherlands)	SIDIAP (Spain)	THIN (UK)	Total
a) Total ever enrolled by 31/12/2015	1,542,672	2,225,925	5,488,397	12,695,046	21,952,040
b) Total adults with ≥1 year enrolment from registration	1,544,573	1,780,500	5,259,575	9,085,325	17,669,973
c) NAFLD patients after exclusion of individuals with a history of alcohol abuse, number (%)	NAFLD: 24,027 (1.56%)	NAFLD: 18,865 (1.06%)	NAFLD: 77,107 (1.47%)	NAFLD: 23,385 (0.26%)	NAFLD: 143,384 (0.81%)
d) NAFLD patients after exclusion because of less than 6 months of follow-up post NAFLD diagnosis, number (%)	NAFLD: 23,131 (1.50%)	NAFLD: 15,669 (incident patients post registration into IPCI database) (0.88%)	NAFLD: 71,672 (1.36%)	NAFLD: 21,039 (0.23%)	NAFLD: 131,511 (0.74%)
e) NAFLD patients after exclusion if less than 6 months of medical history prior to NAFLD diagnosis, number (%)	NAFLD: 22,708 (1.47%)	NAFLD: 13,386 (0.75%)	NAFLD: 69,451 (1.32%)	NAFLD: 20,346 (0.22%)	NAFLD: 125,891 (0.71%)
f) NAFLD patients after exclusion if history of MI or stroke, number (%)	NAFLD: 21,627 (1.40%)	NAFLD: 12,595 (0.71%)	NAFLD: 67,109 (1.28%)	NAFLD: 19,464 (0.21%)	NAFLD: 120,795 (0.68%)
g) Number of matched unexposed patients (ratio unexposed / exposed) after applying all exclusion criteria	Non-NAFLD: 1,707,510 (ratio: 79)	Non-NAFLD: 1,207,378 (ratio: 96)	Non-NAFLD: 4,830,700 (ratio: 72)	1,902,056 (ratio: 98)	Non-NAFLD: 9,647,644

Denominators for all percentages are values in row b): Total adults with ≥1 year enrolment from registration.

Supplementary Table 2 Incidence rate of MI and stroke in four primary care databases

Categories	HSD	IPCI	SIDIAP	THIN	Overall
Acute myocardial infarction, NAFLD patients					
Total number of person-years	124,525	31,426	259,008	85,361	500,320
Number of events	221	137	414	263	1,035
Incidence rate (95%CI) per 1,000 person-years	1.77 (1.55: 2.02)	4.36 (3.66: 5.15)	1.6 (1.45: 1.76)	3.08 (2.72: 3.48)	2.07 (1.94: 2.20)
Acute myocardial infarction, non-NAFLD patients					
Total number of person-years	9,728,567	3,032,175	18,700,000	8,379,073	39,839,815
Number of events	15,014	9,625	23,238	19,946	67,823
Incidence rate (95%CI) per 1,000 person-years	1.54 (1.52: 1.57)	3.17 (3.11: 3.24)	1.24 (1.23: 1.26)	2.38 (2.35: 2.41)	1.70 (1.69: 1.71)
Stroke, NAFLD patients					
Total number of person-years	122,105	31,422	258,006	85,467	497,000
Number of events	962	156	854	215	2,187
Incidence rate (95%CI) per 1,000 person-years	7.88 (7.39: 8.39)	4.96 (4.22: 5.81)	3.31 (3.09: 3.54)	2.52 (2.19: 2.88)	4.40 (4.22: 4.59)
Stroke, non-NAFLD patients					
Total number of person-years	9,586,232	3,030,972	18,700,000	8,393,764	39,710,968
Number of events	60,082	11,902	45,658	16,359	134,001
Incidence rate (95%CI) per 1,000 person-years	6.27 (6.22: 6.32)	3.93 (3.86: 4)	2.45 (2.42: 2.47)	1.95 (1.92: 1.98)	3.37 (3.35: 3.39)

Data presented as incidence rate and their 95% confidence intervals (CI). Overall incidence rates are estimated by dividing the total number of events by the total number of person-years. 95% CI are estimated using an exact Poisson model.

Supplementary Table 3 Description of subsets used in statistical models

Sample subset	Number of patients		Number of MI events		Number of stroke events	
	NAFLD	Matched non-NAFLD	NAFLD	Matched non-NAFLD	NAFLD	Matched non-NAFLD
a) HSD database						
Whole Study Population	21,627	1,707,510	221	15,014	962	60,082
Subset 1	12,647	662,099	126	7,329	719	37,606
b) IPCI database						
Whole Study Population	12,595	1,207,378	137	9,625	156	11,902
Subset 1	6,977	438,582	90	4,704	101	6,059
c) SIDIAP database						
Whole Study Population	67,109	4,830,700	414	23,134	854	45,605
Subset 1	52,188	2,728,743	334	14,877	702	31,539
c) THIN database						
Whole Study Population	19,464	1,902,056	263	19,946	215	16,359
Subset 1	14,286	835,564	197	10,496	144	8,656
d) All databases combined						
Whole Study Population	120,795	9,647,644	1,035	67,719	2,187	133,948
Subset 1	86,098	4,664,988	747	37,406	1,666	83,860

Subset 1 includes individuals with information on total cholesterol, SBP and history of hypertension.

Supplementary Table 4 Descriptive characteristics of patients in the four databases in the whole study population and in subsets with full risk factor data on characteristics in table.

	Characteristics	Matched controls	NAFLD
HSD database	Mean age in years (SD)	60.1 (11.6)	58.8 (12.9)
	Males %	49%	53%
	Current smoker, %	13%	13.40%
	History of Type 2 diabetes, %	17%	22%
	History of hypertension, %	55%	58%
	Statin use, %	24%	25%
IPCI database	Mean age in years (SD)	61.6 (10.7)	59.2 (12.2)
	Males %	45%	47%
	Current smoker, %	17%	20%
	History of Type 2 diabetes, %	20%	32%
	History of hypertension, %	50%	49%
	Statin use, %	39%	45%
SIDIAP database	Mean age in years (SD)	57.9 (12.1)	56.8 (13.0)
	Males %	43%	50%
	Current smoker, %	18%	19%
	History of Type 2 diabetes, %	15%	23%
	History of hypertension, %	40%	47%
	Statin use, %	32%	37%
THIN database	Mean age in years (SD)	58.4 (11.1)	55.2 (12.2)
	Males %	50%	51%
	Current smoker	18%	18%
	History of Type 2 diabetes, %	14%	27%
	History of hypertension, %	45%	49%

Note that the percentage of current smokers is estimated after imputation of patients with missing smoking status as non-current smokers.