Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Incidence and risk factors for cardiovascular diseases in middle-aged and older adults : a 9-year prospective nationwide cohort study in China

Jing Ma
Affiliated Hospital of Chengde Medical University
Jian-Li Wang
Affiliated Hospital of Chengde Medical University
Zhi-Jie Dou
cyfydzj@126.com

Affiliated Hospital of Chengde Medical University

Research Article

Keywords: Cardiovascular diseases, Incidence, Risk factors, CHARLS, Prospective cohort study

Posted Date: June 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9826184/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Identifying modifiable cardiovascular diseases (CVDs) risk factors is crucial for prevention, yet the intricate interactions among lifestyle, clinical, and metabolic factors require analytical approaches that consider mediation and collinearity.

Methods

This study analyzed data from 10,964 participants in the China Health and Retirement Longitudinal Study (CHARLS) from 2011 to 2020. A manual stepwise-forward Cox regression was used, encompassed 59 factors across nine domains, incrementally adding domains to adjust for mediating effects. Sensitivity analyses and subgroup analyses by age and sex were conducted.

Results

During a median follow-up period of 8.26 years, encompassing a total of 54211 person-years, 1566 new CVDs were identified. In the final model, fifteen factors were found to be independently associated with CVDs incidence: dyslipidemia (HR = 1.51, 1.26–1.82), other functional limitations (HR = 1.49, 1.31–1.69), liver disease (HR = 1.49, 1.16–1.91), asthma (HR = 1.47, 1.17–1.85), hypertension (HR = 1.47, 1.26–1.72), kidney disease (HR = 1.34, 1.10–1.64), college or above (HR = 1.31, 1.00–1.72), digestive disease (HR = 1.25, 1.11–1.41), former smoking (HR = 1.23, 1.01–1.51), gender (women) (HR = 1.22, 1.06–1.40), age (HR = 1.01, 1.01–1.02), average diastolic (HR = 1.01, 1.00–1.01), waist (HR = 1.01, 1.00–1.01), platelets (HR = 1.00, 1.00–1.00) and not currently married (HR = 0.86, 0.75–0.98). Sensitivity analyses affirmed the robustness of most associations, with the exception of gender. Furthermore, sex was observed to modify the effect of kidney disease (interaction HR = 1.51, 1.01–2.27).

Conclusions

Multiple lifestyles, chronic disease, physical function, and laboratory examinations are independently associated with CVDs risk. The hierarchical analytical approach revealed that the effects of certain risk factors differ by age and sex, highlighting the value of personalized prevention strategies.

Introduction

Cardiovascular diseases (CVDs), encompassing a range of disorders affecting the heart and blood vessels, represent the leading cause of mortality globally and significantly impair patient quality of life and contribute to disability. According to the World Health Organization, CVDs account for an estimated 17.9 million deaths annually, with more than 80% of these fatalities resulting from heart attacks and

strokes [1]. In China, despite a decline in the age-standardized mortality rate for CVDs from 286.85 per 100,000 in 2005 to 245.39 per 100,000 in 2020, the estimated total number of CVDs related deaths increased from 3.09 million to 4.58 million during the same period [2]. Furthermore, there is a pronounced age-dependent gradient in 10-year CVDs mortality among middle-aged and older adults, with individuals aged 45 years and older constituting a demographic at particularly high risk for contributing to the national CVDs burden [3, 4]. Considering this, it is crucial to prioritize the identification and proactive management of risk factors for CVDs.

Although significant progress has been made in identifying individual risk factors, the complex interplay among lifestyle, clinical, physical, and metabolic determinants (as well as their potential mediating effects) is still not fully understood [5–7]. Traditional analytical approaches, such as multivariate logistic regression, treat all potential variables uniformly and may inadvertently adjust for mediators, thereby diminishing the true causal impacts or obscuring important pathways [3, 8]. Moreover, many earlier studies focus on individual risk factors or not have systematically considered the hierarchical nature of exposures, like lifestyle factors affecting chronic diseases, which affect metabolic markers [9, 10]. This limitation is particularly relevant in CVDs epidemiology, where the long preclinical phase and complex causes demand analytical strategies that respect the sequence of events and differentiate direct from mediated effects [4].

To fill these research gaps, we conducted a thorough prospective cohort study using data from the China Health and Retirement Longitudinal Study (CHARLS), a nationally representative survey focusing on middle-aged and elderly Chinese adults. Our study utilized a novel hierarchical analytical approach that systematically integrated risk factor domains in a fixed order, starting with lifestyle factors and progressing to metabolic markers. This method allowed us to account for potential mediating effects and collinearity between variables. In addition to applying the traditional stepwise Cox regression method, we implemented a manual stepwise forward selection process guided by domain-specific knowledge [11]. We also performed formal tests to assess effect modification by age and sex. Furthermore, we conducted multiple sensitivity analyses to confirm the reliability of our results. This study aims to find independent modifiable risk factors for CVDs and provide a replicable methodological framework for future epidemiological research.

Methods

Study Design and Participants

The CHARLS, initiated in 2011, is an ongoing, nationally representative longitudinal survey conducted in China. This study gathers high-quality data through structured interviews using a standardized questionnaire, targeting a representative sample of the Chinese population aged 45 years and older. The sample is selected using a multistage stratified probability-proportionate-to-size sampling method. The inaugural nationwide survey (Wave 1) was carried out in 2011, encompassing 150 counties and districts across 28 provinces, thereby ensuring extensive geographical and socioeconomic representation.

Subsequent follow-up surveys have been conducted biennially, including Wave 2 (2013–2014), Wave 3 (2015–2016), Wave 4 (2017–2018) and Wave 5 (2019–2020). The dataset includes individual weighting variables to maintain national representativeness. Comprehensive details regarding the CHARLS study design have been documented in previous publications [12]. Informed consent was obtained from all participants, and the study protocol received approval from the Ethical Review Committee of Peking University (approval number: IRB00001052–11,015). All procedures adhered to institutional and national ethical standards, the Declaration of Helsinki (1964, as amended), and equivalent guidelines. This study conformed to the STROBE reporting guideline [13].

The present research was conducted among participants aged over 45 years who did not have a CVDs at baseline (Fig. 1). Initially, 25,586 participants were included at Wave 1. Exclusions were made for individuals younger than 45 years (n = 602), those with CVDs at baseline (n = 2709), those with missing baseline CVDs data (n = 8119), those without a blood sample (n = 7325), and those without follow-up visits (n = 267). Consequently, a total of 6564 eligible participants were enrolled in our analysis (Fig. 1).

Assessment of outcomes

The primary outcome of this study was the incidence of CVDs. During each wave of the CHARLS, participants were asked whether they had ever been diagnosed with heart problem or stroke by a physician. Additionally, they were queried about their current engagement in any heart problem or stroke related treatments, including traditional Chinese medicine, modern Western medicine, chemotherapy, or none of these options. Participants who self-reported a positive diagnosis of heart problem or stroke by a doctor or who indicated undergoing specific treatments for heart problem or stroke were classified as individuals with CVDs.

Risk Factors

All factors available in CHARLS were included for analysis using the baseline data (Data missing more than 20% will be deleted). The following data was collected in this study: (1) demographics: age, gender and education level; (2) sociodemographic: marital status, residence (hukou), currently working, whether self-employed and participate in social activities; (3) lifestyle information: smoking status and alcohol consumption status; (4) body measurements: systolic blood pressure (SBP), diastolic blood pressure (DBP), waist measurement and body mass index (BMI); (5) chronic diseases: hypertension, diabetes, dyslipidemia, lung disease, cancer, arthritis, liver disease, kidney disease, digestive disease and asthma; (6) medication for disease; (7) physical function: activities of daily living (ADLs), instrumental activities of daily living (IADLs), other function limitations, 7 item mobility, upper body mobility, lower body mobility, maximum lung function peak flow and dominant hand grip strength; (8) psychiatric factors: center for epidemiologic studies depression Scale-10 (CESD10), contact children, experienced death of own child and good relationship with female/male guardian; (9) laboratory examinations: white blood cell (WBC), hemoglobin (HB), hematocrit (HCT), mean corpuscular volume (MCV), platelet, C-Reactive Protein (CRP),

glycosylated hemoglobin (HbA1c), fasting blood glucose (FBG), triglyceride (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-c), low-density lipoprotein cholesterol (LDL-c) and blood urea nitrogen (BUN).

Statistical Analysis

Baseline characteristics were separately summarized for individuals with and without CVDs, presented as percentages for categorical variables and as means with standard deviations for normally distributed continuous variables. CVDs incidence rates were examined in 9-year age groups starting at age 45, the earliest recruitment age, and separately for men and women. Cox regression models were employed to assess the relationship between risk factors and new CVDs cases. Age was used as the timeline, beginning at birth, with participants entering the risk pool at their initial assessments (wave 1) and continuing until the first occurrence of CVDs, loss to follow-up, death, or the end of follow-up (wave 5).

In univariate analysis, each candidate factor was analyzed individually by Cox proportional hazards regression models, and adjusted for age, sex, and education level. Variables with a P value less than 0.1 were considered potentially linked to the outcome and were kept for further selection. Then, a stepwise Cox regression analysis was carried out, using Akaike Information Criterion (AIC) as the selection criterion. We employed both forward and backward selection methods. The initial model incorporated all variables that achieved statistical significance ($P < 0.1$) in the univariate analysis. Variables were added or removed iteratively according to AIC until the optimal model was found. The final model was developed when no further refinements in the AIC scores were possible. The results of the stepwise regression were used to identify independent risk factors, which are presented as hazard ratios (HRs) with 95% confidence intervals (CIs).

To improve the selection process, we considered potential dependencies and mediating effects among variables, and used a manual stepwise forward approach. A correlation matrix was first calculated to assess the interrelationships between candidate variables in each domain. All 59 candidate factors were individually analyzed with Cox proportional hazards models, adjusted for age, sex, and education level (Model 2). Variables demonstrating statistical significance at $P < 0.05$ in Model 2 were subsequently incorporated into domain-specific multivariable models, where all significant factors within the same domain were analyzed collectively, with adjustments for age and sex (Model 3). Factors that maintained significance ($P < 0.05$) in these domain-specific models were then advanced to cross-domain analyses. In this part of the study, we created a bunch of sequential models by step adding domains in a fixed order. The process started with the inclusion of sociodemographic factors, followed by the incorporation of lifestyle information in Model 4, body measurements in Model 5, chronic diseases in Model 6, treatment or medication for disease in Model 7, physical function in Model 8, psychiatric variables in Model 9, and blood markers in Model 10. This process ended up with a whole model (Model 11) that included all the variables that were statistically significant in various domains. Model 11 serves as the foundation for our final conclusions. This hierarchical stepwise method allowed us to handle the vast

number of potential predictors effectively, while accounting for possible correlations and mediating influences among the variables.

All statistical analyses were performed with R software version 4.3.1 (http://www.).R-project.org/), and the P-value is considered less than 0.05 as a sign of statistical significance.

Sensitivity and subgroup analyses

We performed three different sensitivity analyses. We first established a composite endpoint consisting of incident CVDs or all-cause mortality, then re-ran the Cox proportional hazards models to incorporate this new definition. Secondly, to prevent potential reverse causation, we removed participants who had developed the outcome within the first two years of follow-up, and re-estimated the models accordingly. Thirdly, we did a worst-case scenario analysis, where loss to follow-up was treated as an event, specifically, participants who lost to follow-up were considered to have experienced the outcome at their last recorded follow-up time.

Moreover, we did subgroup and interaction analyses to check if associations differ by age and sex. Participants were stratified into two groups based on baseline age: younger (< 65 years) and older (≥ 65 years). Separate Cox models were fitted for each age stratum, and interaction terms between age group (dichotomized as < 65 vs. ≥65 years) and each risk factor of interest were incorporated into the comprehensive model to formally test for effect modification.

Results

Baseline characteristics of participants

At baseline, 25586 participants were assessed. After exclusion, 6564 participants were included in the analysis. Baseline characteristics are provided in Table 1. Of 6564 included participants, 3174 (48.4%) were men, and the mean (SD) age was 59.52 (9.80) years. During a total follow-up of 54211 person-years (mean [SD] follow-up duration 8.26 [1.78] years; median [IQR] follow-up 9 years), 1566 incident cases of CVDs were identified. The 9-year cumulative incidence was 23.86%, and the incidence density was 28.89 per 1000 person-years.

Table 1
Baseline Characteristics of individuals.

Characteristics	Overall	Without CVDs	With CVDs	p
n	6564	4998	1566	
Age, years	59.52 ± 9.80	59.11 ± 9.95	60.85 ± 9.16	< 0.001
Gender (Male) (%)	3174 (48.4)	2482 (49.7)	692 (44.2)	< 0.001
Education level (%)				
Elementary school or below	4652 (70.9)	3510 (70.2)	1142 (72.9)	0.011
Middle school	1699 (25.9)	1335 (26.7)	364 (23.2)	
College or above	213 (3.2)	153 (3.1)	60 (3.8)	
Current married (%)	5378 (81.9)	4085 (81.7)	1293 (82.6)	0.477
Residence (Countryside) (%)	1048 (16.0)	797 (15.9)	251 (16.0)	0.97
Never work (%)	2157 (32.9)	1571 (31.4)	586 (37.4)	< 0.001
Self-employed (%)	446 (6.8)	355 (7.1)	91 (5.8)	0.086
Social activities (%)	3032 (46.2)	2320 (46.4)	712 (45.5)	0.528
Smoking status (%)				
Never	4048 (61.7)	3057 (61.2)	991 (63.3)	< 0.001
Former	516 (7.9)	365 (7.3)	151 (9.6)	
Current	2000 (30.5)	1576 (31.5)	424 (27.1)	
Drinking status (%)				
Never	3923 (59.8)	2937 (58.8)	986 (63.0)	< 0.001
Former	527 (8.0)	384 (7.7)	143 (9.1)	
Current	2114 (32.2)	1677 (33.6)	437 (27.9)	
BMI, kg/m2	23.60 ± 13.80	23.42 ± 13.59	24.17 ± 14.43	0.058
Waist measurement (cm)	83.80 ± 11.74	83.29 ± 11.38	85.44 ± 12.68	< 0.001
SBP, mmHg	129.58 ± 20.39	128.42 ± 19.97	133.29 ± 21.24	< 0.001

Characteristics	Overall	Without CVDs	With CVDs	p
DBP, mmHg	75.30 ± 11.54	74.78 ± 11.36	76.96 ± 11.95	< 0.001
Dyslipidemia (%)	483 (7.4)	285 (5.7)	198 (12.6)	< 0.001
Hypertension (%)	1512 (23.0)	963 (19.3)	549 (35.1)	< 0.001
Diabetes (%)	336 (5.1)	224 (4.5)	112 (7.2)	< 0.001
Kidney disease (%)	321 (4.9)	218 (4.4)	103 (6.6)	0.001
Liver disease (%)	188 (2.9)	123 (2.5)	65 (4.2)	0.001
Lunge disease (%)	596 (9.1)	402 (8.0)	194 (12.4)	< 0.001
Digestive disease (%)	1360 (20.7)	972 (19.4)	388 (24.8)	< 0.001
Arthritis (%)	2178 (33.2)	1578 (31.6)	600 (38.3)	< 0.001
Asthma (%)	268 (4.1)	162 (3.2)	106 (6.8)	< 0.001
CVDs (%)	47 (0.7)	38 (0.8)	9 (0.6)	0.556
Medicine for dyslipidemia (%)	208 (3.2)	117 (2.3)	91 (5.8)	< 0.001
Medicine for hypertension (%)	1044 (15.9)	648 (13.0)	396 (25.3)	< 0.001
Medicine for diabetes (%)	201 (3.1)	129 (2.6)	72 (4.6)	< 0.001
Medicine for CVDs (%)	24 (0.4)	20 (0.4)	4 (0.3)	0.556
Medicine for lunge disease (%)	402 (6.1)	262 (5.2)	140 (8.9)	< 0.001
Medicine for Arthritis (%)	992 (15.1)	695 (13.9)	297 (19.0)	< 0.001
Medicine for Liver disease (%)	94 (1.4)	63 (1.3)	31 (2.0)	0.049
Medicine for Kidney disease (%)	180 (2.7)	131 (2.6)	49 (3.1)	0.324
Medicine for Digestive disease (%)	892 (13.6)	649 (13.0)	243 (15.5)	0.012
ADLs (%)	984 (15.0)	674 (13.5)	310 (19.8)	< 0.001

Characteristics	Overall	Without CVDs	With CVDs	p
IADLs (%)	1273 (19.4)	893 (17.9)	380 (24.3)	< 0.001
Other function (%)	4340 (66.1)	3138 (62.8)	1202 (76.8)	< 0.001
7 item mobility	1.17 ± 1.47	1.06 ± 1.41	1.50 ± 1.59	< 0.001
Lower body mobility	0.94 ± 1.12	0.86 ± 1.09	1.20 ± 1.19	< 0.001
Upper body mobility	0.23 ± 0.57	0.20 ± 0.53	0.30 ± 0.66	< 0.001
Lung function peak flow	290.69 ± 118.79	295.19 ± 120.48	276.33 ± 112.05	< 0.001
Dominant hand grip strength (kg)	31.20 ± 10.58)	31.55 ± 10.60)	30.11 ± 10.46	< 0.001
CESD10 Score > 10 (%)	2340 (35.6)	1667 (33.4)	673 (43.0)	< 0.001
Weekly contact children (%)	6062 (92.4)	4608 (92.2)	1454 (92.8)	0.429
Experienced own child death (%)	636 (9.7)	471 (9.4)	165 (10.5)	0.211
Good relationship with female guardian (%)				
Excellent	2071 (31.6)	1522 (30.5)	549 (35.1)	0.006
Very good	2126 (32.4)	1645 (32.9)	481 (30.7)	
Good	1187 (18.1)	935 (18.7)	252 (16.1)	
Fair	1121 (17.1)	850 (17.0)	271 (17.3)	
Poor	59 (0.9)	46 (0.9)	13 (0.8)	
Good relationship with male guardian (%)				
Excellent	1965 (29.9)	1467 (29.4)	498 (31.8)	0.155
Very good	2077 (31.6)	1602 (32.1)	475 (30.3)	
Good	1232 (18.8)	960 (19.2)	272 (17.4)	
Fair	1213 (18.5)	913 (18.3)	300 (19.2)	
Poor	77 (1.2)	56 (1.1)	21 (1.3)	
HbA1c, %	5.25 ± 0.78	5.22 ± 0.75	5.32 ± 0.89	< 0.001

Characteristics	Overall	Without CVDs	With CVDs	p
FBG, mg/dL	109.33 ± 33.93	108.29 ± 31.69	112.66 ± 40.08	< 0.001
TG, mg/dL	128.93 ± 94.08	126.85 ± 94.44	135.57 ± 92.64	0.001
TC, mg/dL	193.06 ± 38.00	191.89 ± 37.63	196.79 ± 38.93	< 0.001
HDL-c, mg/dL	51.62 ± 15.23	51.90 ± 15.26	50.74 ± 15.12	0.008
LDL-c, mg/dL	116.28 ± 34.74	115.29 ± 34.19	119.44 ± 36.25	< 0.001
BUN, mg/dL	15.87 ± 4.64	15.91 ± 4.67	15.76 ± 4.56	0.26
CRP, mg/L	2.62 ± 6.52	2.58 ± 6.45	2.74 ± 6.74	0.404
WBC (in thousands)	6.23 ± 1.89	6.21 ± 1.90	6.30 ± 1.86	0.097
Platelets, 10^9/L	210.53 ± 72.81	209.27 ± 70.84	214.56 ± 78.64	0.012
HB, g/dL	14.37 ± 2.24	14.36 ± 2.24	14.41 ± 2.25	0.524
HCT, %	41.68 ± 6.03	41.65 ± 6.03	41.77 ± 6.04	0.487
MCV	90.72 ± 8.51	90.65 ± 8.52	90.95 ± 8.47	0.235

ADLs: activities of daily living; IADLs: instrumental activities of daily living; CESD: center for epidemiologic studies depression scale; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HbA1c, glycosylated hemoglobin A1c; FBG, fasting blood glucose; TG, triglyceride; TC, total cholesterol; HDL-c, high-density lipoprotein cholesterol; LDL-c, low-density lipoprotein cholesterol; BUN, blood urea nitrogen; CRP, C-reactive protein; WBC, white blood cell; HB, hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume.

The 9-year cumulative incidence increased steadily from the 40–49 years age group to the 60–69 years group for both sexes, then decreased slightly for the older age groups (Fig. 2). In all age bands from 40–49 to 60–69 years, women had higher cumulative incidence than men, with the largest absolute difference observed in the 60–69 years group (women: 31.4%; men: 26.5%). In the 70–79 years group, the cumulative incidence was 23.6% in men and 30.8% in women, still higher in women. However, in the oldest age band (80–89 years), the cumulative incidence was 24.7% in men and 22.1% in women, slightly higher in men. Overall, the 9-year cumulative incidence rises with age, peaks around 60–69 years, and remains substantial in older adults. Women experience a higher cumulative burden than men during middle age, whereas the sex difference narrows and may reverse at very old ages. These findings highlight the importance of age- and sex-specific in the CVDs prevention strategies.

Risk Factors of CVDs

To assess the robustness of our findings, we first performed a stepwise Cox regression using the AIC. This analysis identified 20 factors that were independently associated with CVDs incidence, as detailed in Fig. 3. To further refine the selection while accounting for potential dependencies and mediating effects among variables, we then applied a manual stepwise-forward approach.

CESD: center for epidemiologic studies depression scale; LDL, low-density lipoprotein cholesterol.

Among the 59 factors univariately analyzed in model 1, 49 showed a significant association with CVDs incidence (Fig. 4; Supplement Table 1). These factors were subsequently combined in joint analyses within and across domains. For the domain-specific analyses, correlation matrices were assessed and revealed no highly correlated variables. In the final model derived from this approach, 15 factors retained a significant association with CVDs incidence.

Based on the manual stepwise-forward Cox regression, we identified independent risk factors for CVDs across prespecified domains. This hierarchical approach allowed us to account for potential mediation and interdependencies among variables. In the lifestyle domain, former smoking was the only factor that remained significant throughout all models, with a stable HR of approximately 1.23 (95% CI 1.01–1.51) in the final model. In the chronic disease domain, several factors showed persistent associations. Dyslipidemia and hypertension remained significant in all cross-domain models, with final HRs of 1.51 (95% CI 1.26–1.82) and 1.47 (95% CI 1.26–1.72), respectively. Liver disease, asthma, kidney disease and digestive disease also remained significant, with final HRs of 1.49 (95% CI 1.16–1.91), 1.47 (95% CI 1.17–1.85), 1.34 (95% CI 1.10–1.64) and 1.25 (95% CI 1.11–1.41), respectively. Other chronic conditions (e.g., diabetes, arthritis, lunge disease) were not retained beyond the domain-specific models. Within the physical function domain, other functional limitations were consistently associated with CVDs risk across all models, yielding a final HR of 1.49 (95% CI 1.16–1.91). For metabolic domain, platelets were retained in the final model, although its effect size is mode, with CIs narrowly. The final model (Model 11) also included age, education, gender, not currently married, average diastolic and waist as independent risk factors for CVDs. These factors demonstrated stable effect estimates across the sequential modeling steps, supporting the robustness of their associations.

Sensitivity Analyses

As a sensitivity analysis, we evaluated the reliability of our results under a variety of alternative conditions. The main model was re-estimated after: (i) redefining the outcome as a composite of CVDs incidence or all-cause death; (ii) excluding participants who developed the outcome within the first two years of follow-up to mitigate potential reverse causation; and (iii) applying a worst-case scenario in which loss to follow-up was treated as an event. For each scenario, we took the HRs and 95% CIs for the same set of candidate predictors, so we can directly compare them with the primary analysis.

In various sensitivity models, the direction and strength of the observed associations showed high consistency for the majority of the factors analyzed (Fig. 5). For instance, female was associated with a 13% higher CVDs risk (HR ~ 1.13, 95% CI 0.96–1.34) across all four models. Higher education (college or above) consistently indicated a 35–37% increased risk (HRs 1.35–1.37), with CIs that remained above unity. Former smoking had a stable HR of approximately 1.25–1.28, with narrow CIs excluding the null. Dyslipidemia, hypertension, kidney disease, liver disease, digestive disease, asthma, and functional limitation all exhibited HRs between 1.25 and 1.52 across all sensitivity models, with CIs that consistently excluded 1.0. Notably, the association for hypertension was already significant in the main model (HR 1.42, 95% CI 1.21–1.67) and remained virtually unchanged in the alternative scenarios, adding to the robustness of the finding.

For continuous variables – age, waist circumference, diastolic blood pressure, and platelet count – the effect sizes were very small (HR ≈ 1.00–1.01) and did not meaningfully vary across models, confirming that these associations are stable albeit clinically modest. Importantly, no variable showed a reversal of direction or a substantial change in point estimate, and all CIs remained similar across the four analytic strategies.

In summary, the sensitivity analyses strongly support the robustness of the primary results. The vast majority of risk factors retained their statistical significance and effect magnitude regardless of outcome definition, exclusion of early events, or worst-case imputation of loss to follow-up. For the 15 factors presented here, the evidence is highly consistent and reinforces the main conclusions. These findings confirm that the observed associations are not driven by specific analytical choices and are unlikely to be explained by reverse causation or informative censoring.

Subgroup analyses

Based on the age stratified analysis and interaction tests, the following findings were observed (Fig. 6). In the younger group (< 65 years), significant associations with CVDs risk were observed for dyslipidemia (HR = 1.37, 95% CI 1.10–1.71), hypertension (HR = 1.51, 95% CI 1.25–1.83), kidney disease (HR = 1.37, 95% CI 1.08–1.74), liver disease (HR = 1.47, 95% CI 1.09–1.98), digestive disease (HR = 1.18, 95% CI 1.02–1.36), and other functional limitation (HR = 1.54, 95% CI 1.33–1.78). In the older group (≥ 65 years), female sex (HR = 1.38, 95% CI 1.09–1.75), college education (HR = 1.69, 95% CI 1.13–2.55), former smoking (HR = 1.44, 95% CI 1.07–1.94), dyslipidemia (HR = 1.69, 95% CI 1.20–2.37), hypertension (HR = 1.34, 95% CI 1.02–1.76), asthma (HR = 1.66, 95% CI 1.19–2.31), digestive disease (HR = 1.37, 95% CI 1.11–1.69), and other functional limitation (HR = 1.37, 95% CI 1.06–1.78) were associated with increased risk. Formal tests for interaction with age (dichotomised at 65 years) revealed significant effect modification only for age itself (interaction HR = 0.95, 95% CI 0.93–0.97), indicating that the effect of advancing age on CVDs risk differs substantially between the two age strata. For all other examined variables, no significant age interactions were observed, as the CIs for the interaction terms included unity. These results suggest that several risk factors are independently associated with CVDs in both age groups, but the effect of age itself is modified by the age stratum. The absence of interaction for

lifestyle, metabolic, and most chronic disease factors supports the robustness of their effects irrespective of age.

Based on the sex-stratified Cox regression and interaction analyses, the following findings are summarized (Fig. 7). In men, college or above education (HR = 1.51, 95% CI 1.10–2.06), dyslipidemia (HR = 1.68, 95% CI 1.27–2.22), hypertension (HR = 1.61, 95% CI 1.27–2.03), arthritis (HR = 1.20, 95% CI 1.02–1.41), asthma (HR = 1.46, 95% CI 1.07–2.00), digestive disease (HR = 1.36, 95% CI 1.13–1.63), and other functional limitation (HR = 1.55, 95% CI 1.30–1.85) were all significantly associated with increased CVDs risk. In women, former smoking (HR = 1.77, 95% CI 1.16–2.70), dyslipidemia (HR = 1.42, 95% CI 1.10–1.83), hypertension (HR = 1.37, 95% CI 1.11–1.69), kidney disease (HR = 1.64, 95% CI 1.25–2.14), liver disease (HR = 1.58, 95% CI 1.12–2.22), asthma (HR = 1.47, 95% CI 1.05–2.05), digestive disease (HR = 1.20, 95% CI 1.02–1.40), and other functional limitation (HR = 1.42, 95% CI 1.18–1.70) were significantly associated with increased risk, whereas not currently married (HR = 0.76, 95% CI 0.64–0.91) showed a protective effect. No other examined factors reached statistical significance in either sex. Formal tests for interaction between sex (women vs. men) revealed a significant effect modification for kidney disease: the interaction term yielded an HR of 1.51 (95% CI 1.01–2.27), indicating that the association between kidney disease and CVDs risk is stronger in women than in men. For all other variables, the interaction terms were not statistically significant, with CIs that included unity, suggesting that the associations were generally consistent across sexes. In summary, while several risk factors showed sex-specific associations (e.g., college education and arthritis in men; kidney disease and marital status in women), formal interaction tests suggested that only the effect of kidney disease may differ by sex. These findings underscore the importance of considering sex as a potential effect modifier in CVDs risk assessment.

Discussion

Our large-scale prospective cohort study, which included 6564 middle-aged and elderly Chinese adults, utilized a meticulous domain-based hierarchical modeling method to pinpoint a unique array of independent risk factors associated with the onset of CVDs. During a median follow-up period of roughly 9 years, our study found that age, female sex, college education, former smoking, dyslipidemia, hypertension, kidney disease, liver disease, asthma, digestive disease, other functional limitation, and several other factors were each independently associated with a heightened risk of CVDs. The results highlight the complex relationship among metabolic disorders, chronic inflammation, physical function, and lifestyle factors in the development of CVDs, and also point out notable variations in effects based on age and sex.

In China, smoking is a major CVD risk factor, with 60.2% of men actively smoking and 51.3% of non-smoking women exposed to secondhand smoke at home [14]. A study shows a nonlinear link between cigarette smoke and CVD mortality, indicating even low exposure significantly raises risk, highlighting the impact of modest smoking [15]. In our study, former smoking (HR = 1.23, 95% CI 1.01–1.51) was a persistent risk factor, whereas current smoking did not retain significance after full adjustment. This

pattern is likely explained by the “sick-quitter” effect, where individuals with pre-existing chronic conditions reduce or quit smoking, thereby attenuating the apparent risk [16].

After rigorous adjustment for sociodemographic, lifestyle, body measurement, chronic disease, medication, physical function, psychiatric, and metabolic domains, our study identified dyslipidemia, hypertension, liver disease, kidney disease, digestive disease, and asthma as independent risk factors for CVDs. The persistence of dyslipidemia (HR = 1.51, 95% CI 1.26–1.82) and hypertension (HR = 1.47, 95% CI 1.26–1.72) aligns with landmark studies showing that these are among the most important modifiable risk factors for myocardial infarction and stroke [17, 18], with hypertension being the single largest contributor to global CVDs burden [19]. The link between liver disease and increased CVDs risk (HR = 1.49, 95% CI 1.16–1.91) is strongly supported by meta-analyses demonstrating that non-alcoholic fatty liver disease (NAFLD) independently predicts cardiovascular events [20]. Similarly, kidney disease (HR = 1.34, 95% CI 1.10–1.64) is a well-established risk factor, with reduced eGFR and albuminuria each independently associated with CVDs [21]; interestingly, we found a stronger association in women (interaction HR = 1.51), suggesting sex-specific vulnerability. The association of digestive disease (HR = 1.25, 95% CI 1.11–1.41) with CVDs risk is consistent with studies linking inflammatory bowel disease to atherosclerosis, possibly mediated by chronic inflammation and gut dysbiosis [22]. Finally, asthma (HR = 1.47, 95% CI 1.17–1.85) has been shown to increase CVDs risk through systemic inflammation and oxidative stress [23]. The robustness of these associations after extensive adjustment underscores the multifactorial nature of CVDs and supports a holistic, domain-based approach to risk assessment.

Our study introduces a unique and innovative finding: "other functional limitations" have been identified as an independent risk factor for CVDs. This factor demonstrates a significant impact, with a HR of approximately 1.49, and its influence becomes even more pronounced in younger adults, reaching an HR of around 1.54. Although functional limitations are generally considered consequences of chronic diseases, our results indicate that they might also act as a robust biomarker for subclinical physiological decline (frailty) that occurs prior to a CVDs diagnosis. This finding aligns with recent studies demonstrating that frailty is associated with a higher risk of incident cardiovascular events, including myocardial infarction and stroke [24]. A meta-analysis of prospective cohort studies reported that frail individuals had a significantly increased risk of CVDs (HR = 1.70, 95% CI 1.18–2.45) and pre-frail (HR = 1.23, 95% CI 1.07–1.36) vs. robust groups [25]. These findings align with the concept of "accelerated aging," where individuals with lower physical function may accumulate cellular damage, immune surveillance is compromised, and have a pro-inflammatory state, which increases their risk of CVDs [26].

The observed association between platelets reached statistical significance, yet the magnitude of these effects was minimal (HR ~ 1.00). Although those association remained consistent across multiple sensitivity analyses, their relevance to clinical practice is constrained [27]. From a clinical standpoint, those small effect size is usually considered to have limited predictive value for individuals. In the context of population health, that metabolic marker is significant not just by their individual effects, but because they are prevalent in the general population [28]. This concept is reflected in Rose's prevention

paradox, which suggests that a preventive measure that provides little benefit to individuals can still lead to significant public health improvements when implemented widely [29].

Strengths and Limitations

Our results align with the existing, substantial body of evidence indicating that smoke, abnormal lipid levels, high blood pressure, and chronic inflammatory diseases are associated with a heightened CVDs risk. But this study makes several significant advances to existing knowledge. By using a manual stepwise-forward method that sequentially added domains in a fixed order—starting with lifestyle and moving to metabolic factors—we successfully distinguished direct effects from those possibly influenced by later clinical conditions [11]. This method differs from traditional stepwise selection techniques, which consider all variables equally and might unintentionally account for mediating factors, thus reducing the ability to identify true causal relationships. A lot of factors, like hypertension and dyslipidemia, were still important even after accounting for later-stage variables, which suggests they directly contribute to CVDs risk. Comparing to other factors like waist and BMI, these factors did not remain, maybe because their effects are mediated by chronic diseases or metabolic disturbances, which are accounted for later in the hierarchy. Secondly, this study stands out as one of the limited number of investigations that simultaneously evaluates risk factors across nine different domains within a Chinese population, employing a nationally representative sample. There a lot of prior studies focusing on individual risk factors or a few domains, but our whole approach gives a more comprehensive view of the multifactorial etiology of CVDs. In each domain, the application of correlation matrices showed that multicollinearity wasn't a big problem, so our multivariate models are more valid.

Several limitations are to be considered in our study. Even after making extensive adjustments for a wide range of covariates, residual confounding may still remain. This stems from the presence of unmeasured or inaccurately measured confounding factors, including genetic predispositions, specific dietary patterns, or cumulative environmental exposures, which could have impacted the observed correlations. Secondly, all risk factors were evaluated only at the baseline. This time point evaluation is singular, it doesn't account for changes within individual over the follow-up period, which might result in misclassification of exposure and attenuation of true effect estimates, especially regarding modifiable factors like lipid levels or blood pressure. Thirdly, the study population was formed from a single cohort, which might limit the generalizability of our findings to populations with diverse ethnic, socioeconomic, or geographic backgrounds. The internal validity of the study might be strong, but the external validity, especially the transportability of effect estimates, needs to be evaluated in future replication studies. So, a lot of point estimates, like those related to waist, average diastolic and platelets, were close to the null value, with CIs narrowly covering unity. Although these associations were statistically significant, it's probably because of a large sample size, which makes their effect sizes small, so we need to be cautious. Such small effects might be more susceptible to bias due to unmeasured confounding variables or measurement errors.

Conclusions

This study gives a whole lot of and a strong evaluation of independent risk factors linked to CVDs. Using a hierarchical analytical framework, we found fifteen factors that consistently predict CVDs in various sensitivity analyses. Most of these factors are potentially modifiable, and they align with established CVDs risk and protective factors, so they are important in prevention efforts. The observed effects of age and sex further highlight the potential for targeted, subgroup-specific prevention strategies. If validated externally, these findings can be incorporated into current CVDs prevention strategies. Although there is more research need to understand the underlying mechanisms, addressing these modifiable factors might be a good strategy to reduce CVDs risk, and it can be easily added into current prevention initiatives.

Declarations

Acknowledgments

This study utilize data from the CHARLS. The authors thank the CHARLS staff team for their efforts and all participants involved for contributing the data.

Authors' contribution

MJ was responsible for conceptualizing and designing the study, extracting and collating the data, and conducting the statistical analysis and interpretation. JLW drafted the manuscript and conducted the literature search. DZJ contributed to manuscript revisions, supervised the research and provided critical review. All authors reviewed and approved the submitted version of the manuscript.

Funding

This work was supported by grants from Clinical Medicine Excellent Talent Training Project in Hebei Province (Grant No. ZF2023245) and Chengde Science and Technology Agency (Grant No. 202109A193).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the China Health and Retirement Longitudinal Study repository [http:// Charls. pku. edu. cn].

Clinical trial number: Not applicable.

Ethics approval and consent to participate:

The study was approved by the ethics review committee (institutional review board) of Peking University (approval number: IRB00001052–11,015). All participants granted formal written consent for their involvement.

Consent to Participate: Not applicable.

Consent for Publication:

All authors have given final approval of the version and agreed with the publication of this study here.

Conflict of Interest Statement:

The authors declare that they have no conflict of interest.

References

1. Camacho-Encina M, Booth LK, Redgrave RE, Folaranmi O, Spyridopoulos I, Richardson GD. Cellular Senescence, Mitochondrial Dysfunction, and Their Link to Cardiovascular Disease. Cells 2024, 13(4).
2. Wang W, Liu Y, Liu J, Yin P, Wang L, Qi J, You J, Lin L, Meng S, Wang F, et al. Mortality and years of life lost of cardiovascular diseases in China, 2005–2020: Empirical evidence from national mortality surveillance system. Int J Cardiol. 2021;340:105–12.
3. Huang Q, Jiang Z, Shi B, Meng J, Shu L, Hu F, Mi J. Characterisation of cardiovascular disease (CVD) incidence and machine learning risk prediction in middle-aged and elderly populations: data from the China health and retirement longitudinal study (CHARLS). BMC Public Health. 2025;25(1):518.
4. Clarke R, Halsey J, Emberson J, Collins R, Leon DA, Kivimäki M, Shipley MJ. Lifetime and 10-year risks of cardiovascular mortality in relation to risk factors in middle and old age: 50-year follow-up of the Whitehall study of London Civil Servants. Public Health. 2024;230:73–80.
5. Center For Cardiovascular Diseases The Writing Committee Of The Report On, Cardiovascular H. Diseases In China N: Report on Cardiovascular Health and Diseases in China 2023: An Updated Summary. Biomed Environ Sci. 2024;37(9):949–92.
6. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Das SR, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. Circulation. 2019;139(10):e56–528.
7. Conrad N, Molenberghs G, Verbeke G, Zaccardi F, Lawson C, Friday JM, Su H, Jhund PS, Sattar N, Rahimi K, et al. Trends in cardiovascular disease incidence among 22 million people in the UK over 20 years: population based study. BMJ. 2024;385:e078523.
8. Wu J, Wang Y, Li S, He H, Xu Y, Yang Q, Shen H. Epidemiological trends, risk factor contributions, and future projections of cardiovascular disease among Chinese adults ≥ 45 years: insights from GBD 2021 and CHARLS. Exp Gerontol. 2026;216:113072.
9. Jiang M, Ren X, Han L, Zheng X. Associations between sarcopenic obesity and risk of cardiovascular disease: A population-based cohort study among middle-aged and older adults using the CHARLS. Clin Nutr. 2024;43(3):796–802.
10. You J, Guo Y, Kang JJ, Wang HF, Yang M, Feng JF, Yu JT, Cheng W. Development of machine learning-based models to predict 10-year risk of cardiovascular disease: a prospective cohort study. Stroke Vasc Neurol. 2023;8(6):475–85.

11. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology. 2009;20(4):488–95.
12. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol. 2012;43(1):61–8.
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344–9.
14. Gu D, Wu X, Reynolds K, Duan X, Xin X, Reynolds RF, Whelton PK, He J. Cigarette smoking and exposure to environmental tobacco smoke in China: the international collaborative study of cardiovascular disease in Asia. Am J Public Health. 2004;94(11):1972–6.
15. Pope CA 3rd, Burnett RT, Turner MC, Cohen A, Krewski D, Jerrett M, Gapstur SM, Thun MJ. Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure-response relationships. Environ Health Perspect. 2011;119(11):1616–21.
16. Sarich P, Gao S, Zhu Y, Canfell K, Weber MF. The association between alcohol consumption and all-cause mortality: An umbrella review of systematic reviews using lifetime abstainers or low-volume drinkers as a reference group. Addiction. 2024;119(6):998–1012.
17. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364(9438):937–52.
18. O'Donnell MJ, Chin SL, Rangarajan S, Xavier D, Liu L, Zhang H, Rao-Melacini P, Zhang X, Pais P, Agapay S, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet. 2016;388(10046):761–75.
19. Global burden. of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1223–49.
20. Targher G, Byrne CD, Lonardo A, Zoppini G, Barbui C. Non-alcoholic fatty liver disease and risk of incident cardiovascular disease: A meta-analysis. J Hepatol. 2016;65(3):589–600.
21. Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, de Jong PE, Coresh J, Gansevoort RT. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet. 2010;375(9731):2073–81.
22. Feng W, Chen G, Cai D, Zhao S, Cheng J, Shen H. Inflammatory Bowel Disease and Risk of Ischemic Heart Disease: An Updated Meta-Analysis of Cohort Studies. J Am Heart Assoc 2017, 6(8).
23. Jiang Y, Huang X, Yu D, Xu C, Wang Y, Wang X, Shen Y. Asthma and the risk of cardiovascular diseases and mortality: a meta-analysis of cohort studies. Ther Adv Respir Dis. 2025;19:17534666251333965.

24. He D, Wang Z, Li J, Yu K, He Y, He X, Liu Y, Li Y, Fu R, Zhou D, et al. Changes in frailty and incident cardiovascular disease in three prospective cohorts. Eur Heart J. 2024;45(12):1058–68.
25. Veronese N, Cereda E, Stubbs B, Solmi M, Luchini C, Manzato E, Sergi G, Manu P, Harris T, Fontana L, et al. Risk of cardiovascular disease morbidity and mortality in frail and pre-frail older adults: Results from a meta-analysis and exploratory meta-regression analysis. Ageing Res Rev. 2017;35:63–73.
26. Ferrucci L, Fabbri E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. Nat Rev Cardiol. 2018;15(9):505–22.
27. Kaiser R, Nicolai L. Procoagulant platelets: linking coagulation and thromboinflammation in cardiovascular disease. Nat Rev Cardiol 2026.
28. Peacock JL, Coto SD, Rees JR, Sauzet O, Jensen ET, Fichorova R, Dunlop AL, Paneth N, Padula A, Woodruff T, et al. Do small effects matter more in vulnerable populations? an investigation using Environmental influences on Child Health Outcomes (ECHO) cohorts. BMC Public Health. 2024;24(1):2655.
29. Rose G. Sick individuals and sick populations. Int J Epidemiol. 1985;14(1):32–8.

Figures

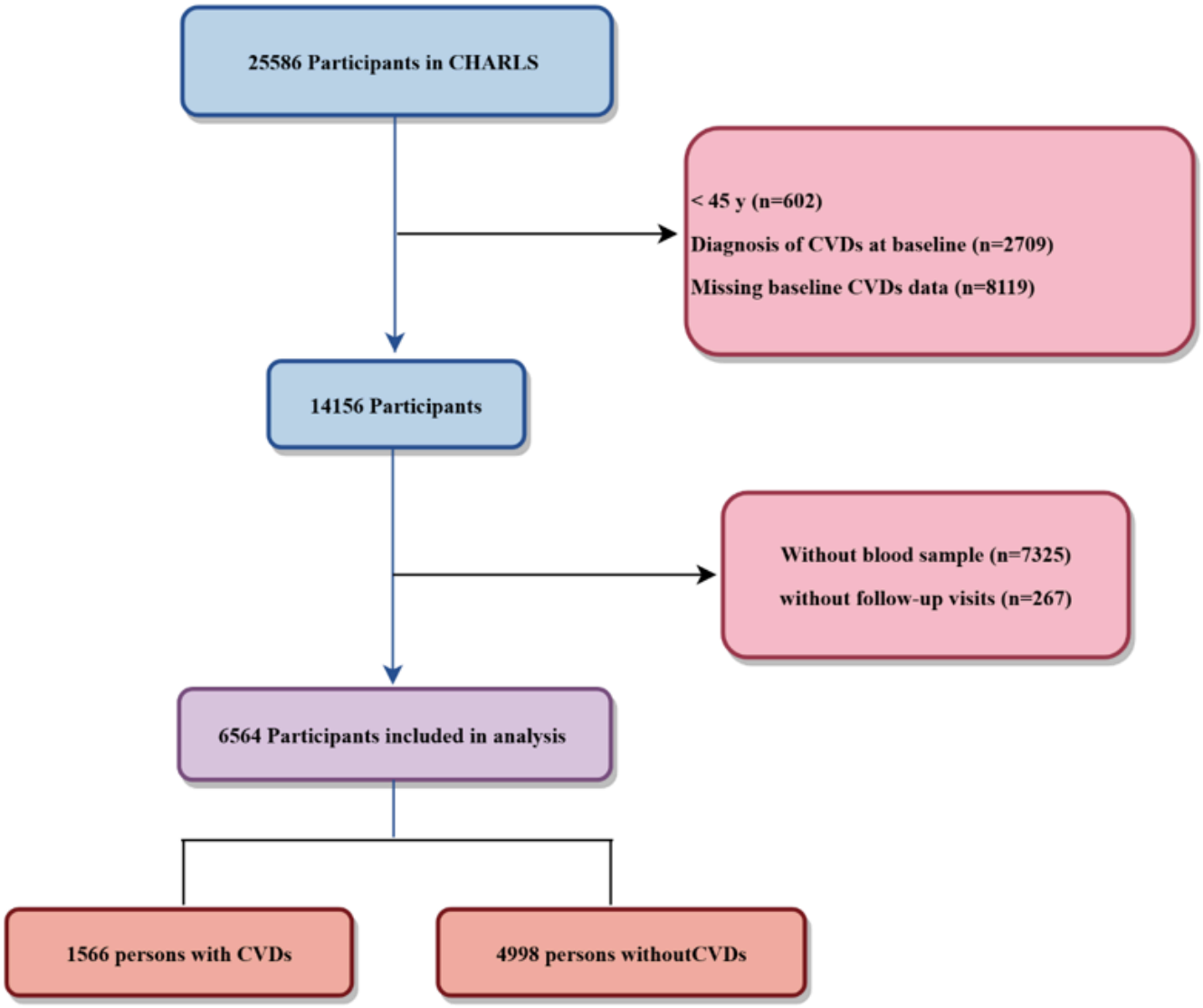


Figure 1

Flowchart of Included Participants.

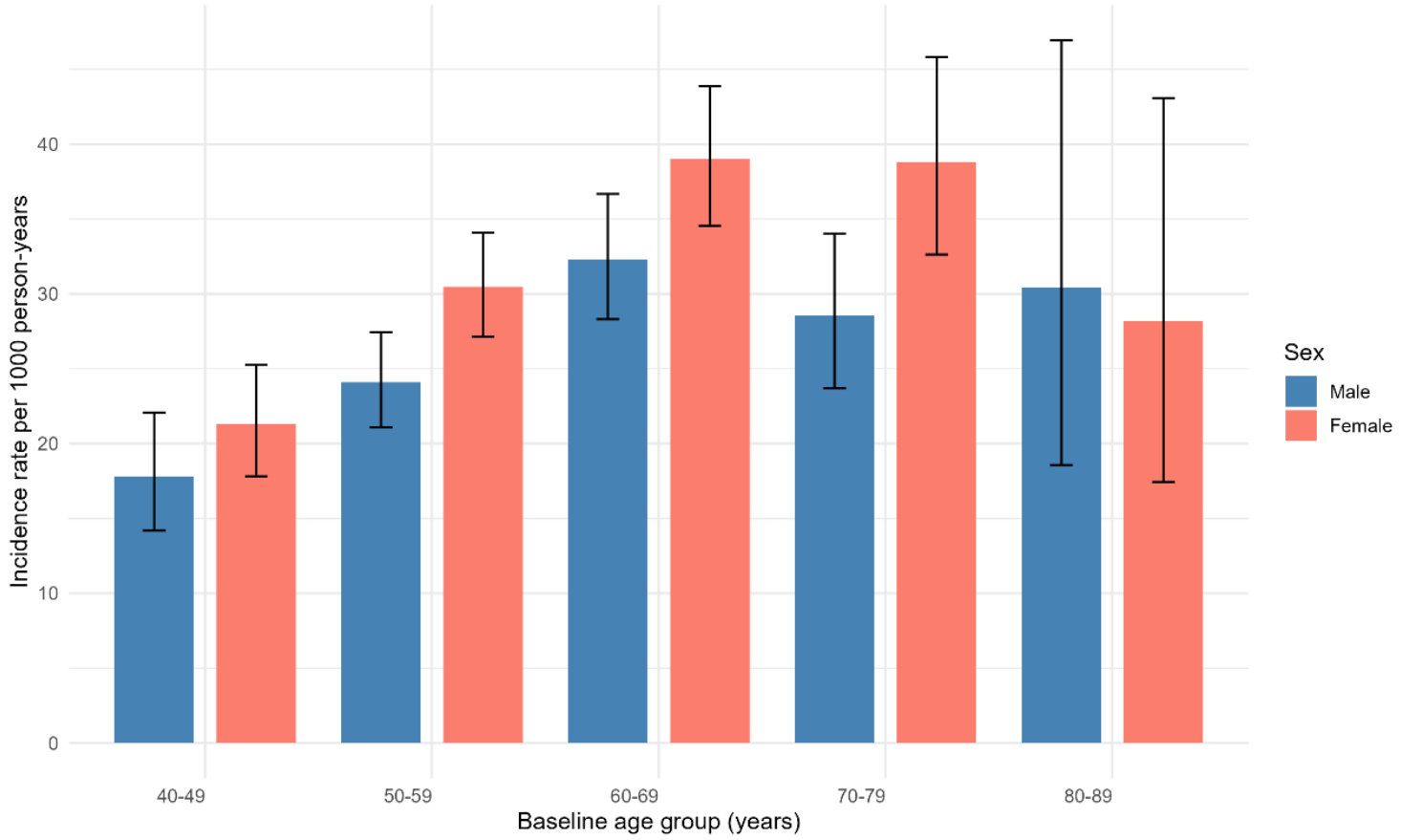


Figure 2

Incidence Rates of CVDs in the CHARLS by 10-Year Age Bands and Sex.

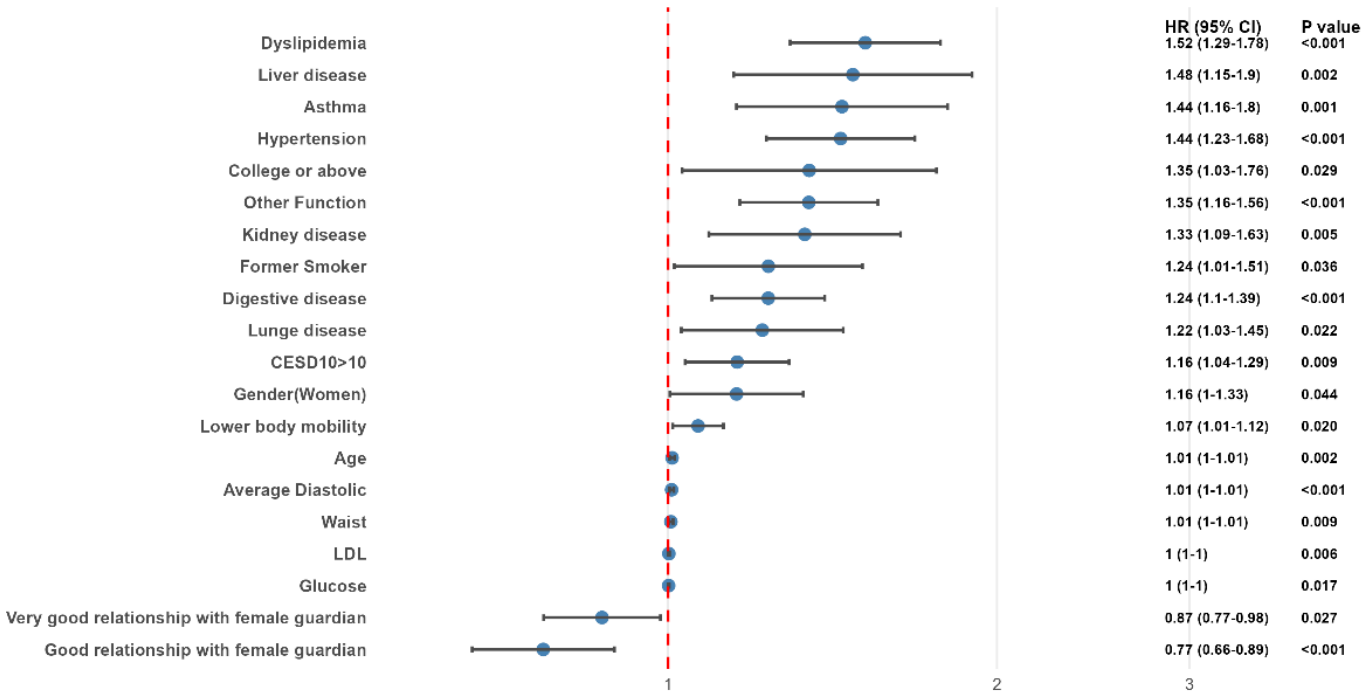


Figure 3

Risk Factors Associated with CVDs using a stepwise Cox regression.

CESD: center for epidemiologic studies depression scale; LDL, low-density lipoprotein cholesterol.

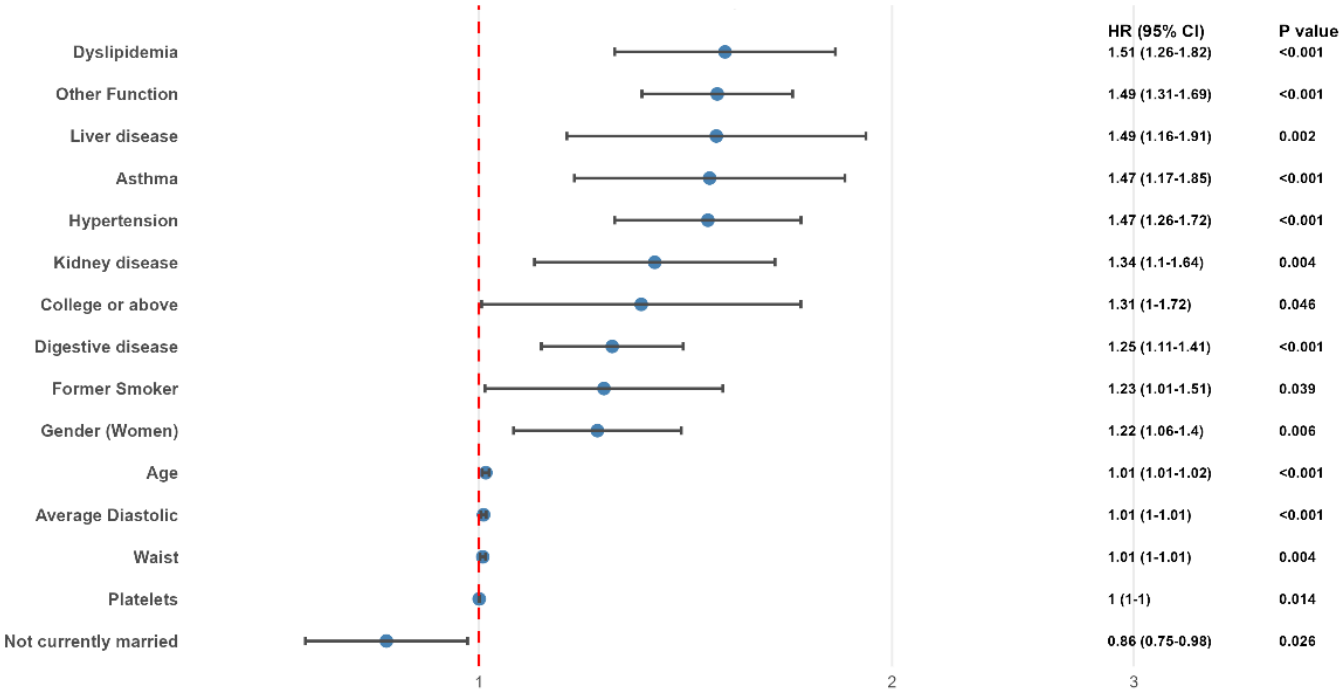


Figure 4

Risk Factors Associated with CVDs in the fully adjusted Model.

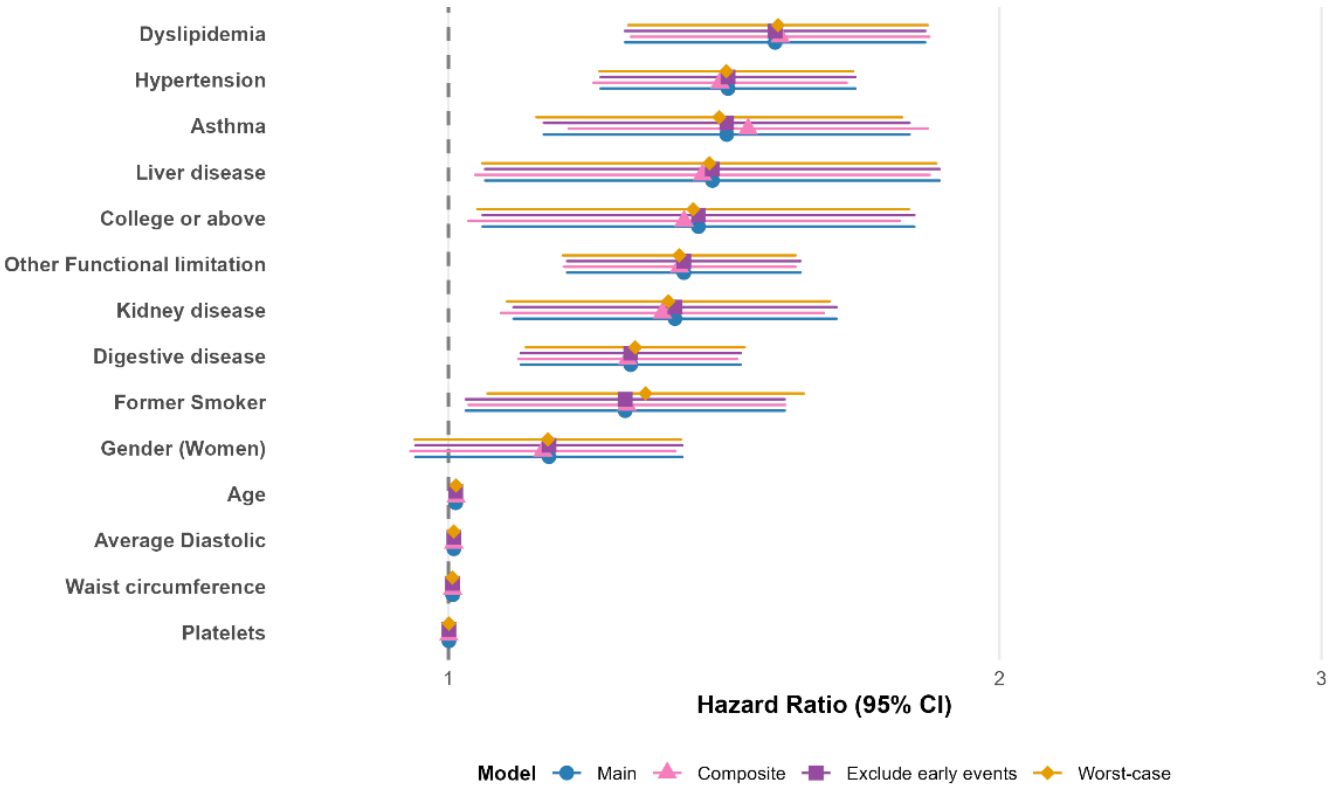


Figure 5

Sensitivity analysis of risk factors.

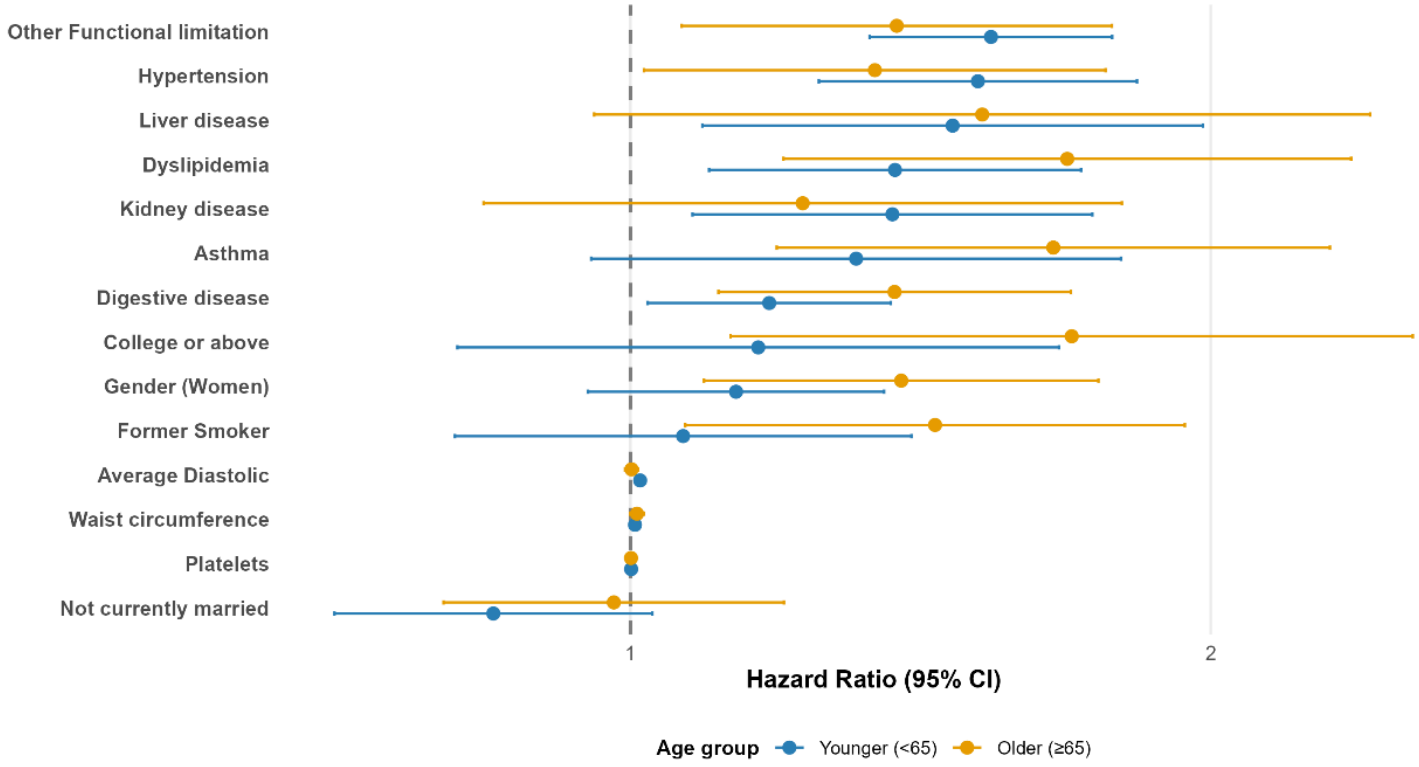


Figure 6

Age-stratified Hazard Ratios for CVDs Risk.

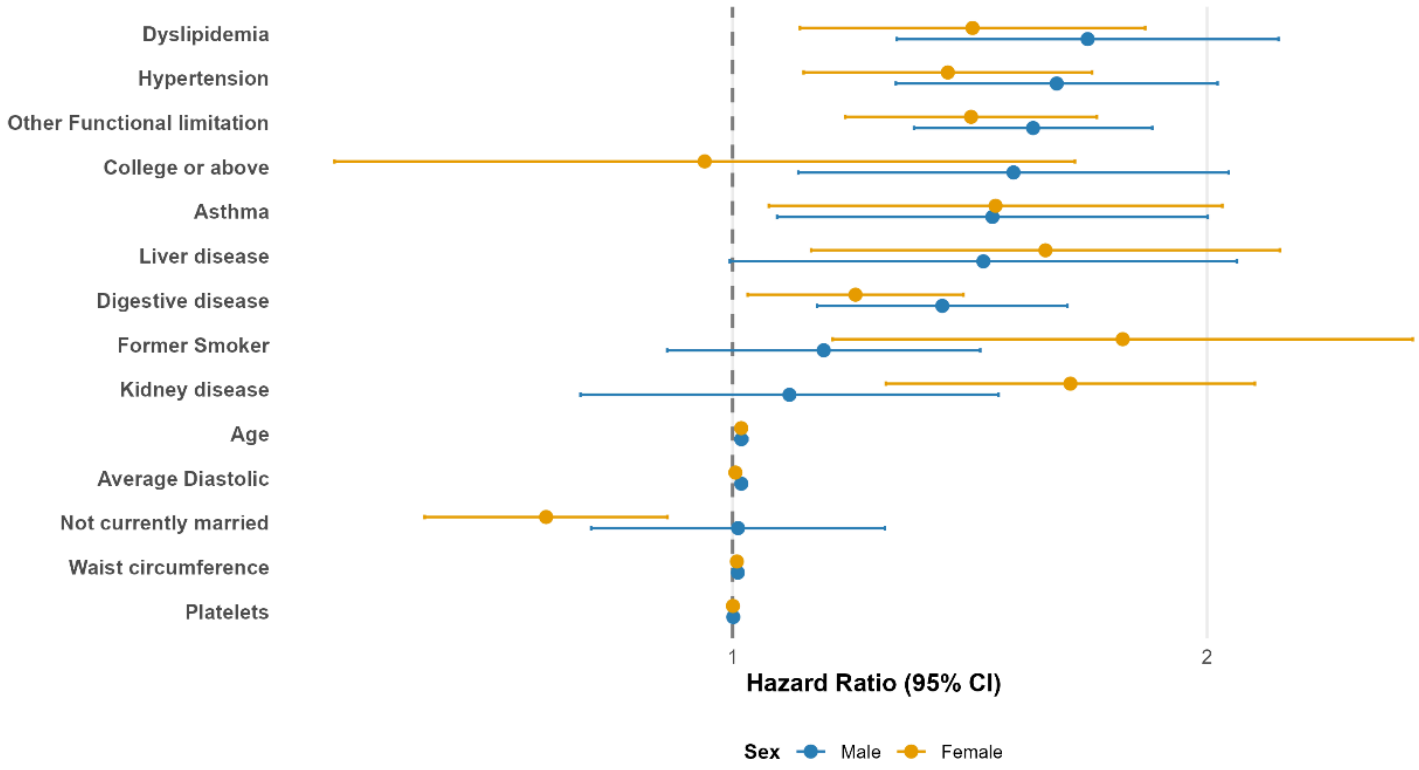


Figure 7

Sex-stratified Hazard Ratios for CVDs risk.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementTable.docx