

Metabolic risk factors, genetic predisposition, and risk of severe liver disease in Chinese: a prospective study of 0.5 million people

Short title: Metabolic risk factors, genetics and liver disease

Yuanjie Pang, DPhil¹, Jun Lv, PhD¹, Christiana Kartsonaki, DPhil^{2,3}, Canqing Yu, PhD¹, Yu Guo, MSc⁴, Yiping Chen, DPhil^{2,3}, Ling Yang, PhD^{2,3}, Iona Y. Millwood, DPhil^{2,3}, Robin G. Walters, PhD^{2,3}, Sisi Wang, BM⁵, Junshi Chen, MD⁶, Zhengming Chen, DPhil^{2,3}, Liming Li, MD¹

1. Department of Epidemiology & Biostatistics, School of Public Health, Peking University, 38 Xueyuan Road, Beijing 100191, China
2. Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, Big Data Institute Building, Roosevelt Drive, University of Oxford, UK
3. Medical Research Council Population Health Research Unit (MRC PHRU) at the University of Oxford, Nuffield Department of Population Health, University of Oxford, UK
4. Chinese Academy of Medical Sciences, 9 Dongdan San Tiao, Beijing 100730, China
5. Liuzhou Center for Disease Prevention and Control, 1 Tanzhongxi Road, Liuzhou 545007, China
6. National Center for Food Safety Risk Assessment, 37 Guangqu Road, Beijing 100021, China

Address for correspondence:

Prof. Liming Li
Department of Epidemiology and Biostatistics
School of Public Health
Peking University
38 Xueyuan Road
Beijing, 100191, China
Tel: 86-010-82801528
Fax: 86-010-82801530
lmleeph@vip.163.com

Word count: abstract 270, main text 5185

(3 tables, 2 figures and 7 supplementary tables/figures)

Revised on February 2021

Sources of support

Dr Pang acknowledges support from the China Postdoctoral Science Foundation (2019TQ0008, 2020M670071). This work was supported by grants (2016YFC0900500, 2016YFC0900501, 2016YFC0900504, 2016YFC1303904) from the National Key R&D Program of China. The CKB baseline survey and the first re-survey were supported by a grant from the Kadoorie Charitable Foundation in Hong Kong. The long-term follow-up is supported by grants from National Natural Science Foundation of China (91846303, 91843302, 81390540, 81390541, 81390544), and Chinese Ministry of Science and Technology (2011BAI09B01). The funders had no role in the study design, data collection, data analysis and interpretation, writing of the report, or the decision to submit the article for publication.

Data share statement

Data described in the article, code book, and analytic code will be made available from CKB upon request (<https://www.ckbiobank.org/site/Data+Access>) or from SCHS upon request pending application and approval. The data sets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Abbreviations list

ALT, alanine aminotransferase; AST, aspartate transaminase; BMI, body mass index; CI, confidence interval; CKB, China Kadoorie Biobank; DSP, Disease Surveillance Points; FLI, fatty liver index; FPG, fasting plasma glucose; GGT, gamma-glutamyl transferase; GRS, genetic risk score; GWAS, genome-wide association study; HBV, hepatitis B virus; HC, hip

circumference; HR, hazards ratio; ICD-10, International Classification of Diseases, 10th Revision; MET, metabolic equivalent of task; NAFLD, non-alcoholic fatty liver disease; PAF, population attributable fractions; RPG, random plasma glucose; SBP, systolic blood pressure; SD, standard deviation; SLD, severe liver disease; WC, waist circumference; WHR, waist-to-hip ratio.

1 **Abstract**

2 **Background:** Metabolic risk factors have been shown to be associated with severe liver
3 disease (SLD) in Chinese. However, there is limited evidence on the combined impact of
4 these factors or the genetic variants associated with SLD.

5 **Objective:** We examined the associations of combined metabolic risk factors with risk of
6 SLD, both overall and by genetic predisposition to SLD.

7 **Design:** The study population involved 486,828 participants of the prospective China
8 Kadoorie Biobank aged 30-79 years from 10 diverse areas in China without history of cancer
9 or liver disease at baseline. Cox regression was used to estimate adjusted hazard ratios
10 (HRs) for SLD associated with combined metabolic risk factors (central adiposity, physical
11 inactivity, and diabetes) by stratum of genetic risk (assessed separately by a *PNPLA3* variant
12 [rs738409] and a body mass index [BMI] genetic risk score).

13 **Results:** During ~10 years of follow-up, 3279 incident cases of SLD were recorded. The
14 overall mean (standard deviation) BMI was 23.8 (3.4) kg/m², and 5.9% participants had
15 diabetes. Compared with those with 3 metabolic factors, participants with 2, 1, and 0
16 metabolic factors had a 31% (HR 0.69 [95% CI 0.65, 0.73]), 43% (0.57 [0.53, 0.60]), and
17 52% (0.48 [0.42, 0.56]) lower risk of SLD, respectively. For both BMI and NAFLD variants,
18 participants with fewer metabolic factors had a lower risk of SLD and lower levels of gamma-
19 glutamyl transferase and fatty liver index in participants with low and high genetic risk (*p*-
20 value for interaction >0.05).

21 **Conclusions:** In relatively lean Chinese adults, individuals with fewer metabolic risk factors
22 had a lower relative risk of SLD and a more favorable profile of liver biomarkers across all

23 strata of genetic risk.

24 **Keywords:** metabolic risk factors; gene-environment interaction; liver cancer; cirrhosis

25 **Introduction**

26 Severe liver disease (SLD), including liver cancer and cirrhosis, has caused substantial
27 morbidity and mortality worldwide (1, 2). According to the Global Burden of Disease project,
28 there were 2.1 million deaths from SLD in 2017, with 20% of these deaths occurred in China
29 (1). SLD has a major impact in China, with hepatitis B virus (HBV) being the most important
30 risk factor. In China, more than 50% of SLD cases were attributable to HBV in 1990, and the
31 proportion was 40% in 2013 (3). With the increasing prevalence of a sedentary lifestyle, a
32 growing proportion of SLD is attributable to metabolic risk factors in recent decades (3, 4).

33 Prospective studies in Western populations have reported several metabolic risk factors for
34 SLD, including body mass index (BMI), leisure-time physical activity, plasma glucose, and
35 diabetes (5-8). Recent prospective studies have confirmed these associations in Chinese
36 and expanded the phenotypes of metabolic risk factors and disease subtypes (9-12). For
37 example, a 2019 study showed that central adiposity was more important than general
38 adiposity in predicting risk of chronic liver disease and liver cancer (9, 10), while a 2020
39 study reported an inverse association of total physical activity with chronic liver disease
40 regardless of disease subtypes (11). However, the combined impact of these metabolic risk
41 factors on SLD risk has yet to be quantified. Such information is useful as these factors are
42 associated with SLD through shared mechanisms involving the insulin resistance syndrome
43 (13) and are potentially modifiable through lifestyle interventions.

44 Apart from modifiable risk factors, the genetics of SLD have been extensively investigated
45 (14-16). Although no genetic risk score (GRS) has been developed for SLD to date, genetic
46 predispositions to BMI and non-alcoholic fatty liver disease (NAFLD) have been shown to

47 be associated with SLD (17). Previous case-control studies in China have confirmed
48 *PNPLA3* p.I148M (rs738409) as a risk allele for NAFLD (18), while a 2020 study reported a
49 positive genetic association between BMI GRS (consisting of 97 alleles) and risk of chronic
50 liver disease in Chinese (19). However, it is unclear whether genetic factors can modify the
51 associations between combined metabolic risk factors and risk of SLD. Assessing the
52 interactions between modifiable metabolic risk factors and genetic factors on SLD would
53 provide insights into risk stratification and targeted interventions.

54 Therefore, we aimed to examine (1) the associations of combined metabolic risk factors with
55 risk of SLD based on recommendations of three potentially modifiable factors (central
56 adiposity, physical inactivity, and hyperglycemia/diabetes); and (2) whether the associations
57 of combined metabolic risk factors with risk of SLD differed according to genetic
58 predisposition to SLD (assessed by BMI GRS and rs738409). To explore the underlying
59 mechanisms, we also assessed the associations of combined metabolic risk factors with
60 liver biomarkers (liver enzymes and fatty liver index [FLI]) by genetic predisposition to SLD.

61 **Subjects and Methods**

62 *Study population*

63 Details of the CKB design, survey methods and population characteristics have been
64 described elsewhere (20). Briefly, 512,715 participants (210,205 men and 302,510 women)
65 aged 30-79 years were recruited into the study from 10 (5 urban, 5 rural) geographically
66 defined localities in China during 2004-2008. The study areas were selected to provide
67 diversity in risk exposure and disease patterns, while taking into account population stability,
68 quality of mortality and morbidity registries, capacity, and long-term commitment within the

69 areas. Prior international, national and regional ethical approvals were obtained, and all
70 participants provided written informed consent. At local study assessment clinics,
71 participants completed an interviewer-administered laptop-based questionnaire on socio-
72 demographic characteristics, smoking, alcohol consumption, diet, physical activity, personal
73 and family medical history and current medication. A range of physical measurements were
74 recorded by trained technicians, including height, weight, hip and waist circumference, bio-
75 impedance, lung function, blood pressure and heart rate, using calibrated instruments with
76 standard protocols.

77 *Data collection on metabolic risk factors*

78 Standing height was measured using a stadiometer. Weight was measured using a body
79 composition analyzer (TANITA-TBF-300GS; Tanita Corporation), with subtraction of weight
80 of clothing according to season (ranging from 0.5 kg in summer to 2.0-2.5 kg in winter).
81 Waist circumference (WC) and hip circumference (HC) were measured using a soft non-
82 stretchable tape, with HC measured at the maximum circumference around the buttocks.
83 BMI at baseline was calculated as the measured weight in kilograms divided by the square
84 of the measured height in meters. WHR is the ratio of WC to HC.

85 A 10-ml non-fasting (with the time since the participant last ate recorded) blood sample was
86 collected from participants, with immediate on-site testing of random plasma glucose (RPG)
87 level using the SureStep Plus System (Johnson & Johnson), regularly calibrated with
88 manufacturer quality control solution. Participants with glucose levels ≥ 7.8 mmol/L and < 11.1
89 mmol/L were invited to return for a fasting plasma glucose (FPG) test the next day. RPG
90 data were unavailable for 8156 (1.6%) participants (because of a delay in making the on-

91 site test available in certain regions). Previously diagnosed diabetes was defined by a
92 positive response to the question “Has a doctor ever told you that you had diabetes?”.
93 Among those without previously diagnosed diabetes, screen-detected diabetes was defined
94 as RPG ≥ 7.0 mmol/L and time since last eating ≥ 8 h, or ≥ 11.1 mmol/L with time since last
95 eating < 8 h, or a FPG ≥ 7.0 mmol/L on subsequent testing (21). Diabetes was defined as
96 having either previously diagnosed diabetes or screen-detected diabetes.

97 Participants were asked about their usual type and duration of activities related to work,
98 commuting, household chores, and leisure-time exercise during the past year. To quantify
99 the amount of physical activity, metabolic equivalent of tasks (METs) from the 2011 update
100 of a major compendium of physical activities were used (22). The MET value for a particular
101 type of physical activity represents the ratio of the energy expended per kilogram of body
102 weight per hour during that activity to that expended when sitting quietly. The number of
103 hours spent per day participating in each activity was multiplied by the MET score for that
104 activity, and the daily amount of total physical activity was obtained by summing the MET-
105 hours for activities related to occupation and non-occupational (i.e. commuting, housework,
106 and non-sedentary leisure-time activities) activities.

107 *Follow-up for and ascertainment of disease cases*

108 The vital status of each participant was determined periodically through China CDC’s
109 Disease Surveillance Points (DSP) system and national health insurance system (which
110 covers over 98% of our participants, capturing all episodes of hospitalization, for
111 reimbursement purposes), supplemented by regular checks against local residential and
112 administrative records and by annual active confirmation through street committees or

113 village administrators (23). In addition, information about major diseases and any episodes
114 of hospitalization was collected through linkages, via each participant's unique national
115 identification number, with disease registries (for cancer, ischemic heart disease, stroke, and
116 diabetes) and national health insurance claims databases. All disease events were coded
117 using International Classification of Diseases, 10th Revision (ICD-10) by trained DSP staff
118 (for death) or medical professionals (for hospitalized events) who were blinded to baseline
119 information (10). SLD was defined as ICD-10 code K74 (cirrhosis) and C22 (liver cancer).
120 In CKB, C22 was used only for primary liver cancer and secondary cancer was coded as
121 'C78.7'. By January 1, 2017, 44,066 (9%) participants had died and 4751 (<1%) were lost
122 to follow-up.

123 *Genotyping and biochemistry measurements*

124 Genotyping was conducted using a custom-designed 800K-SNP array (Axiom; Affymetrix)
125 with imputation to 1000 Genomes Phase 3. Genotype data were available for samples from
126 100,408 participants passing QC (overall call rate >99.97% across all variants). This
127 included a population-based sample of 75,736 participants randomly selected from the total
128 CKB cohort, and a further 24,672 participants who had been selected for nested case-
129 control studies of incident stroke, coronary heart disease or chronic obstructive pulmonary
130 disease (**Supplementary Figure 1**). To avoid potential selection bias, only the 75,736
131 randomly selected participants were used for genetic analyses of hepatobiliary outcomes.
132 17,567 genotyped participants who had been selected for nested case-control studies of
133 stroke and of coronary heart disease were assayed for liver function biomarkers (alanine
134 aminotransferase [ALT], aspartate transaminase [AST], and gamma-glutamyl transferase

135 [GGT]) at the Wolfson Laboratory (Oxford) using baseline plasma samples. Further details
136 of genotyping and biochemistry are in the **Supplementary methods**.

137 *Definition of metabolic risk factors*

138 We selected physical activity, central adiposity, and diabetes to construct a combined
139 metabolic score. This is because these metabolic factors have been shown to be associated
140 with risk of chronic liver disease in the Chinese population (9-12). To investigate the
141 combined effects of metabolic risk factors, we grouped each participant into 1 of 4 categories
142 according to the number of metabolic risk factors (0-3), including central obesity (WHR \geq 0.90
143 [men] or 0.85 [women]), diabetes (self-reported or screen-detected diabetes), and physical
144 inactivity (total physical activity <33.2 MET-h/day [the top 20%]). The cut-off points were
145 selected for each metabolic factor based on *a priori* knowledge of the risk factors for chronic
146 liver disease and are considered achievable at the population level (9-12).

147 *Definition of genetic factors*

148 To investigate whether the association of the combined metabolic factors and risk of SLD
149 differed by genetic predisposition, we considered two genetic factors (i.e. BMI and NAFLD
150 variants). For BMI, a genetic risk score was built based on 92 risk variants identified in
151 previous genome-wide association studies (24), which was also confirmed in CKB (19). The
152 score was calculated as the sum of risk alleles of the respective BMI-increasing variants (0,
153 1, or 2 copies per risk allele). For NAFLD, we used rs738409, a well-established genetic
154 variant for NAFLD in both Caucasians and East Asians. *PNPLA3*p.I148M genotype was
155 coded 0, 1, and 2 for non-carriers, heterozygous, and homozygous of the risk-increasing
156 allele (i.e. the M allele). For each genetic factor, participants were divided into two groups

157 according to the genetic predisposition: low genetic risk (BMI GRS <median or 0 M allele)
158 and high genetic risk (BMI GRS ≥median or 1-2 M alleles).

159 *Statistical analysis*

160 The present study excluded individuals with a prior history of cancer (n=2578), cirrhosis or
161 hepatitis (n=6139), and positive HBsAg tests (n=17,170), leaving 486,828 individuals for the
162 main analysis. The primary outcome were SLDs (composite SLD, cirrhosis, and liver cancer)
163 and the secondary outcomes were liver biomarkers (ALT, AST, GGT, and fatty liver index
164 [FLI]). The formula of FLI is described in **Supplementary methods**. In the analysis of
165 individual metabolic risk factors, Cox proportional hazards regression models were used to
166 estimate adjusted hazard ratios (HRs) for SLD, adjusted for age at baseline, sex, study area
167 (10 regions), education (4 groups: no formal school, primary school, middle/high school, or
168 college/university), smoking (4 groups: never, occasional, former regular, or current regular),
169 alcohol (5 groups: abstainers, occasional drinkers, reduced-intake drinkers, ex-weekly
170 drinkers, or weekly drinkers), BMI, and self-rated health, with additional adjustment for the
171 other metabolic factors. Time since birth was used as the underlying time scale and
172 participants entered the study at their baseline age.

173 In combined analyses, the metabolic factors were modelled as a categorical variable (0 to 3
174 points) and as an ordinal variable (per 1-point decrease in number; linear trend). The same
175 variables were adjusted for as in the analysis of individual risk factors, and each metabolic
176 factor was weighted equally. Adjusted HRs were reported for individuals with 0, 1, and 2
177 metabolic risk factors compared with those with all 3 metabolic risk factors. Combined
178 analyses were also performed separately for cirrhosis and liver cancer. For analyses

179 involving more than two categories, all HRs are presented with 95% CIs calculated using
180 'floating' standard errors to facilitate comparisons between any two groups rather than just
181 with the reference group (25). Instead of selecting one level of the risk factor as the reference
182 group, a 'floated' variance is assigned to each level which describes the uncertainty in risk
183 without reference to another level (25).

184 To investigate potential influences on the association between the combined metabolic
185 factors and risk of SLD, we performed analyses stratified by genetic risk of BMI and NAFLD
186 separately. In addition, we examined the associations of the combination of genetic and
187 metabolic categories (6 categories, with high genetic risk and three metabolic risk factors as
188 reference) with risk of SLD and tested an interaction between metabolic and genetic factors
189 using the likelihood ratio test.

190 In addition, population attributable fractions (PAFs) and 95% CIs were calculated to estimate
191 the proportion of SLD cases that were attributable to the individual metabolic factors as well
192 as combined metabolic risk factors (26, 27). Statistical analyses were done using R 2.14.2.

193 **Results**

194 *Baseline characteristics of study participants*

195 Of the 486,828 participants included, mean (SD) age was 52 (10.9) years, and 59.2% were
196 women. The mean (SD) BMI was 23.8 (3.4) kg/m², and 5.9% had diabetes at baseline.
197 During 10 years of follow-up, there were 3279 incident cases of SLD, including 1558 cases
198 of cirrhosis and 2025 cases of liver cancer. Participants with fewer metabolic risk factors
199 were younger, less likely to be female and from urban areas, and had lower levels of
200 education and income (**Table 1**). Participants with fewer metabolic risk factors had lower

201 levels of SBP, RPG, and adiposity, and they were less likely to have prevalent CVD and
202 hypertension and a family history of diabetes or cancer (**Table 1**). Of the 71,952 participants
203 with genetic data, there were generally no differences in socio-demographics by genetic risk
204 groups (**Supplementary Table 1**). Participants with high BMI GRS were more likely to have
205 general adiposity (BMI and percent body fat) and prevalent diabetes and hypertension, while
206 those with high genetic risk of NAFLD were more likely to have prevalent diabetes
207 (**Supplementary Table 1**).

208 *Associations between combined metabolic risk factors and SLD*

209 In the analysis of individual metabolic factors, each of the three factors was associated with
210 risk of SLD (**Table 2**). Participants with central adiposity had a 20% higher risk of SLD (HR
211 1.20 [1.11, 1.31]); the corresponding risk increment was 23% and 49% for physically inactive
212 individuals and those with diabetes (HR 1.23 [1.11, 1.37] and 1.49 [1.33, 1.68]), respectively.
213 Additional adjustment for the other two metabolic factors did not alter the association with
214 each individual factor (**Table 2**). When examining the dose-response associations, there
215 was a positive association for each of the three metabolic factors with risk of SLD
216 (**Supplementary Table 2**).

217 In the analysis of combined metabolic risk factors, participants with fewer metabolic factors
218 had a lower risk of SLD (**Table 3**). Compared with those with 3 metabolic factors, participants
219 with 2, 1, and 0 metabolic factors had a 31% (HR 0.69 [0.65, 0.73]), 43% (0.57 [0.53, 0.60]),
220 and 52% (0.48 [0.42, 0.56]) lower risk of SLD, respectively. Each 1-point decrease in the
221 number of combined metabolic factors was associated with a 20% lower risk of SLD (HR
222 0.80 [0.75, 0.84]). When restricting to participants with genetic data (i.e. the population

subset randomly selected from the baseline cohort), similar associations were observed (Table 3). When assessing cirrhosis and liver cancer as separate outcomes, the associations for individual and combined metabolic factors were similar to those observed for SLD (Table 2 and Supplementary Table 2).

Associations between combined metabolic factors and SLD by genetic factors

Among 71,952 participants with available genetic data, each 1-SD higher genetically-determined BMI was associated with a 72% higher risk of SLD (HR 1.72 [1.06, 1.80]). Participants carrying 2 M alleles of rs738409 had a 32% higher risk of SLD (HR 1.32 [1.01, 1.71]) compared with those with 0 M allele, while the corresponding HR for those carrying 1 M allele was 0.96 (0.78, 1.17).

For both BMI and NAFLD variants, participants with fewer metabolic factors had a lower risk of SLD, regardless of the genetic risk (Figure 1). For example, among participants with low genetic risk of adiposity, the HR for those with 2, 1, and 0 metabolic factors was 0.73 (0.59, 0.90), 0.66 (0.53, 0.82), and 0.43 (0.24, 0.77), respectively; the corresponding HR among participants with higher genetic risk was 0.57 (0.47, 0.69), 0.37 (0.29, 0.48), and 0.40 (0.24, 0.69), respectively. The likelihood ratio test showed that genetic factors did not modify the associations between the combined metabolic factors and risk of SLD (p -value for interaction: 0.09 for BMI and 0.72 for NAFLD). When examining the HR per 1-point decrease in the number of combined metabolic factors, participants with high genetic risk of adiposity tended to have lower risk of SLD compared with those with low genetic risk (0.69 [0.56, 0.86] and 0.81 [0.66, 1.01]), but the differences were non-significant (p -value for heterogeneity 0.30).

245 Compared with participants with high BMI GRS and 3 metabolic factors, participants with
246 high BMI GRS and 2 metabolic factors had a 43% lower risk of SLD, while those with high
247 BMI GRS and 0-1 metabolic factors had a 59% lower risk (**Supplementary Table 3**).
248 Although participants with low BMI GRS and 3 metabolic factors had a non-significant lower
249 risk of SLD (0.74 [0.45, 1.22]), those with low BMI GRS and 2 metabolic factors had 45%
250 lower risk, while those with low BMI GRS and 0-1 metabolic factors had 56% lower risk
251 (**Supplementary Table 3**). A similar pattern was observed by genetic risk of NAFLD
252 (**Supplementary Table 3**).

253 *Associations between combined metabolic factors and liver biomarkers by genetic factors*

254 For both FLI and GGT, participants with fewer metabolic factors had lower levels of
255 biomarkers regardless of the genetic risk of adiposity and NAFLD (**Figure 2**). For FLI, a 1-
256 point decrease in the number of combined metabolic factors was associated with 0.24 SD
257 lower FLI among participants with lower and higher genetic risk of adiposity (-0.24 [-0.28, -
258 0.20] and -0.24 [-0.28, -0.20]), while a 1-point decrease was associated with greater
259 decrease in FLI among participants with lower genetic risk of NAFLD than those with higher
260 genetic risk (-0.27 [-0.31, -0.23] and -0.22 [-0.26, -0.19], *p*-value for interaction 0.02;
261 **Supplementary Table 4**). For GGT, neither genetic predisposition to adiposity nor NAFLD
262 modified the associations of combined metabolic factors with biomarkers (**Supplementary**
263 **Table 4**). For both AST and ALT, there was an inverse association between the number of
264 combined metabolic factors and biomarker levels among participants with lower BMI GRS
265 but not among those with higher BMI GRS (**Supplementary Figure 2** and **Supplementary**
266 **Table 4**). For rs738409, there were no clear associations in either stratum of genetic risk of

267 NAFLD (**Supplementary Figure 2** and **Supplementary Table 4**).

268 *Population attributable fraction of metabolic risk factors*

269 If the associations of metabolic risk factors with SLD risk are largely causal, then the
270 proportion of all SLD cases attributable to metabolic risk factors would be 6.6% for central
271 adiposity, 3.7% for physical activity, and 23.7% for diabetes (**Supplementary Table 5**).

272 **Discussion**

273 In this Chinese population, individuals with fewer metabolic risk factors had a lower relative
274 risk of SLD regardless of the genetic predisposition to the disease. Each 1 fewer metabolic
275 risk factor was associated with a 20% lower risk of SLD. Individuals with fewer metabolic
276 risk factors also had more favorable liver biomarkers, particularly GGT and FLI, among
277 individuals with low and high genetic risk. The current study extended the literature by
278 quantifying the associations of combined metabolic risk factors with clinical and subclinical
279 outcomes of liver disease and may inform risk stratification using genetic variants in Chinese.
280 The study findings have public health significance as these metabolic risk factors are inter-
281 related and can be targeted through lifestyle interventions.

282 Previous prospective studies in Western populations have examined the associations of
283 metabolic risk factors in relation to SLD (5-8). These studies have suggested that adiposity,
284 physical inactivity, hyperglycemia, and diabetes are associated with higher risk of SLD (5-
285 8). In line with these findings, our study in the Chinese population showed that three major
286 metabolic factors (central adiposity, physical activity, and hyperglycemia/diabetes) were
287 associated with risk of SLD (9-12). Apart from individual metabolic risk factors, previous
288 studies have examined multiple metabolic risk factors in one study (28-30), but none has

289 quantified the combined impact of these factors on risk of SLD. Our study showed a graded
290 decrease in risk of SLD with fewer metabolic risk factors. Furthermore, we assessed a range
291 of liver biomarkers and showed consistent associations of metabolic risk factors with
292 biomarkers. Although more biomarkers on the pathophysiological pathway to SLD are
293 warranted, our study findings provide valuable insights into the underlying mechanisms
294 linking metabolic risk factors and SLD.

295 Previous studies have examined the joint effects of NAFLD SNPs and individual metabolic
296 risk factors on liver-related outcomes, suggesting synergetic effects on liver-related
297 biomarkers (16, 31). A genome-wide association study (GWAS) of 10 thousand participants
298 (95% Europeans) reported synergetic effects of NAFLD SNPs and BMI on liver fat content
299 (31). Specifically, the authors found that the effect of NAFLD variants (including rs738409)
300 on liver fat content was significantly amplified by increasing BMI. Likewise, a recent cohort
301 study showed synergetic effects of NAFLD SNPs and metabolic risk factors on ALT (16).
302 This study involved 10 thousand participants from Denmark and UK and showed that the
303 positive association of NAFLD GRS with ALT was amplified by higher BMI and among
304 individuals with diabetes. With regards to SLD, this study reported no interactions of NAFLD
305 GRS (consisting of rs738409, rs58542926, and rs72613567) and metabolic risk factors on
306 cirrhosis or liver cancer, consistent with findings of the current study. On the other hand, no
307 previous studies have examined gene-environment interactions for liver disease using BMI
308 SNPs. In this Chinese population, both BMI GRS and rs738409 were associated with higher
309 risk of SLD. We found a similar pattern of gene-environment interactions using both NAFLD
310 SNP and BMI GRS, suggesting that both genetic predispositions might be used to stratify
311 patients at risk for SLD and to inform targeted preventions.

312 The strengths of the CKB included a prospective design, a large and diverse study
313 population, detailed adjustment for risk factors for SLD, and the validity of the genetic score
314 developed for BMI. In particular, we measured a range of biomarkers in a nested case-
315 control study, which allowed us to examine the potential mechanism between metabolic risk
316 factors and SLD. Our study also had several limitations. First, liver diseases were
317 ascertained through linkages to death and disease registries as well as health insurance
318 databases but there was no proactive assessment of the liver using ultrasound or non-
319 invasive measures of fibrosis (e.g. Fibrosis Score 4, AST-Platelet Ratio Index) (32). However,
320 by using biomarker data available in CKB and constructing FLI, we showed that the main
321 results for liver biomarkers were generally consistent with those for liver events. Second, the
322 analysis of combined metabolic risk factors and genetic factors were conducted in ~75,000
323 participants with genetic data, with a much smaller sample size than the overall associations
324 for metabolic risk factors. However, we compared the associations of combined metabolic
325 risk factors with SLD risk in all participants and those with genetic data, which showed
326 consistent associations (**Table 2**). Nonetheless, we observed a borderline significant
327 interaction (p -value 0.09) between combined metabolic risk factors and BMI GRS on SLD
328 risk, suggesting that greater risk reduction might be achieved through controlling of
329 metabolic risk factors among participants with high genetic risk of adiposity. Third, the
330 current study included BMI and NAFLD variants to assess the genetic predisposition to SLD.
331 Therefore, it is possible that other components of genetic predisposition are not included.
332 Fourth, ascertainment of physical activity in CKB was based on self-reported information.
333 Therefore, we cannot rule out potential recall bias or misclassification of exposures, which
334 may bias the associations towards the null. Fifth, the cut-off point for physical activity was

335 selected according to the distribution of total physical activity in CKB due to the lack of
336 national recommendations in Chinese. Nonetheless, sensitivity analyses showed that even
337 with a lower cut-off of physical activity, the results would not change materially (HR per 1-
338 point decrease in the number of combined metabolic factors: ≥ 33.2 MET-h/day, 0.80 [0.75,
339 0.84]; ≥ 17.5 MET-h/day [median], 0.83 [0.79, 0.88]). Sixth, although participants with a
340 positive test for HBsAg were excluded from all analyses, this test might have yielded false
341 negative results. Finally, residual confounding due to unmeasured or unknown variables (e.g.
342 subclinical infections) cannot be ruled out.

343 This study in a Chinese population showed that individuals with fewer metabolic risk factors
344 had a lower relative risk of SLD regardless of genetic predisposition to the disease.
345 Individuals with fewer metabolic risk factors had a more favorable profile of liver biomarkers
346 in participants with low and high genetic risk. The European Association for the Study of the
347 Liver (EASL) guidelines recommend lifestyle modifications (promoting physical activity and
348 weight loss) as primary prevention strategies for SLD (33), but there have been no
349 guidelines on the primary prevention of liver diseases in China. Our study findings serve as
350 the evidence base to support the role of a healthy metabolic profile in preventing SLD in
351 China. Although future studies are warranted in Chinese to replicate novel signals (e.g.
352 TM6SF2 and HSD17B13) from GWAS conducted in Europeans and to discover new genetic
353 predisposition to metabolic-associated fatty liver disease, genetic risk factors might inform
354 screening strategies and targeted interventions among people at risk of SLD.

355 **Acknowledgment**

356 The chief acknowledgment is to the participants, the project staff, and the China National
357 Centre for Disease Control and Prevention (CDC) and its regional offices for access to death
358 and disease registries. The Chinese National Health Insurance scheme provides electronic
359 linkage to all hospital admission data.

360 LL and ZC had full access to the data. YP, JL, CK, LL and ZC conducted data analysis and
361 are responsible for accuracy of the results and the decision to submit for publication. All
362 authors were involved in study design, conduct, long-term follow-up, review and coding of
363 disease events, interpretation of the results, or writing the report.

364 **Conflicts of interest**

365 None.

366

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1736-1788.
2. Wang FS, Fan JG, Zhang Z, Gao B, Wang HY. The global burden of liver disease: the major impact of China. *Hepatology*. 2014;60(6):2099-2108.
3. Zhou M, Wang H, Zhu J, Chen W, Wang L, Liu S, Li Y, Wang L, Liu Y, Yin P, et al. Cause-specific mortality for 240 causes in China during 1990-2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. *Lancet*. 2016;387(10015):251-272.
4. Global Burden of Disease Liver Cancer Collaboration. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the Global Burden of Disease Study 2015. *JAMA Oncol*. 2017;3(12):1683-1691.
5. Florio AA, Campbell PT, Zhang X, Zeleniuch-Jacquotte A, Wactawski-Wende J, Smith-Warner SA, Sinha R, Simon TG, Sesso HD, et al. Abdominal and gluteofemoral size and risk of liver cancer: The liver cancer pooling project. *Int J Cancer*. 2020;147(3):675-685.
6. Liu B, Balkwill A, Reeves G, Beral V, Million Women Study Collaborators. Body mass index and risk of liver cirrhosis in middle aged UK women: prospective study. *BMJ*. 2010;340:c912.
7. Baumeister SE, Leitzmann MF, Linseisen J, Schlesinger S. Physical activity and the risk of liver cancer: a systematic review and meta-analysis of prospective studies and a bias analysis. *J Natl Cancer Inst*. 2019;111(11):1142-1151.
8. Lofffield E, Freedman ND, Lai GY, Weinstein SJ, McGlynn KA, Taylor PR, Männistö S, Albanes D, Stolzenberg-Solomon RZ. Higher glucose and insulin levels are associated with risk of liver cancer and chronic liver disease mortality among men without a history of diabetes. *Cancer Prev Res (Phila)*. 2016;9(11):866-874.
9. Pang Y, Kartsonaki C, Turnbull I, Guo Y, Chen Y, Clarke R, Bian Z, Bragg F, Millwood IY, Yang L, et al. Adiposity in relation to risks of fatty liver, cirrhosis and liver cancer: a prospective study of 0.5 million Chinese adults. *Sci Rep*. 2019;9(1):785.
10. Pang Y, Kartsonaki C, Guo Y, Chen Y, Yang L, Bian Z, Bragg F, Millwood IY, Yu C, Lv J, et al. Central adiposity in relation to risk of liver cancer in Chinese adults: A prospective study of 0.5 million people. *Int J Cancer*. 2019;145(5):1245-1253.
11. Pang Y, Lv J, Kartsonaki C, Yu C, Guo Y, Du G, Bennett D, Bian Z, Chen Y, Yang L, et al. Association of physical activity with risk of hepatobiliary diseases in China: a prospective cohort study of 0.5 million people. *Br J Sports Med*. 2020 Aug 21; bjsports-2020-102174. doi:10.1136/bjsports-2020-102174.

12. Pang Y, Kartsonaki C, Turnbull I, Guo Y, Clarke R, Chen Y, Bragg F, Yang L, Bian Z, Millwood IY, et al. Diabetes, plasma glucose, and incidence of fatty liver, cirrhosis, and liver cancer: a prospective study of 0.5 million people. *Hepatology*. 2018;68(4):1308-1318.
13. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*. 2006;444(7121):840-846.
14. Caruso S, O'Brien DR, Cleary SP, Roberts LR, Zucman-Rossi J. Genetics of hepatocellular carcinoma: novel approaches to explore molecular diversity. *Hepatology*. 2020 May 28. doi:10.1002/hep.31394.
15. Pfeiffer RM, Rotman Y, O'Brien TR. Genetic determinants of cirrhosis and hepatocellular carcinoma due to fatty liver disease: what's the score? *Hepatology*. 2020;72(3):794-796.
16. Gellert-Kristensen H, Richardson TG, Davey Smith G, Nordestgaard BG, Tybjaerg-Hansen A, Stender S. Combined effect of PNPLA3, TM6SF2, and HSD17B13 variants on risk of cirrhosis and hepatocellular carcinoma in the general population. *Hepatology*. 2020;72(3):845-856.
17. Carter AR, Borges MC, Benn M, Tybjaerg-Hansen A, Davey Smith G, Nordestgaard BG, Lawlor DA. Combined association of body mass index and alcohol consumption with biomarkers for liver injury and incidence of liver disease: a Mendelian randomization study. *JAMA Netw Open*. 2019;2(3):e190305. doi:10.1001/jamanetworkopen.2019.0305.
18. Zou ZY, Wong VW, Fan JG. Epidemiology of nonalcoholic fatty liver disease in non-obese populations: meta-analytic assessment of prevalence, genetic, metabolic and histological profiles. *J Dig Dis*. 2020 May 5. doi:10.1111/1751-2980.12871.
19. Pang Y, Kartsonaki C, Lv J, Millwood IY, Yu C, Guo Y, Chen Y, Bian Z, Yang L, Chen J, et al. Observational and genetic associations of body mass index and hepatobiliary diseases in a relatively lean Chinese population. *JAMA Netw Open*. 2020;3(10):e2018721. doi:10.1001/jamanetworkopen.2020.18721.
20. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, Li L, China Kadoorie Biobank (CKB) collaborative group. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40(6):1652-1666.
21. Bragg F, Li L, Bennett D, Guo Y, Lewington S, Bian Z, Yang L, Chen J, Chen Y, Collins R, et al. Association of random plasma glucose levels with the risk for cardiovascular disease among Chinese adults without known diabetes. *JAMA Cardiol*. 2016;1(7):813-823.
22. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, Greer JL, Vezina J, Whitt-Glover MC, Leon AS. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43(8):1575-1581.

23. Yang G, Rao C, Ma J, Wang J, Wan X, Dubrovsky G, Lopez AD. Validation of verbal autopsy procedures for adult deaths in China. *Int J Epidemiol*. 2006;35(3):741-748.
24. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206.
25. Easton DF, Peto J, Babiker AG. Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. *Stat Med*. 1991;10(7):1025-1035.
26. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol*. 1985;122(5):904-914.
27. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol*. 1974;99(5):325-332.
28. Jarvis H, Craig D, Barker R, Spiers G, Stow D, Anstee QM, Hanratty B. Metabolic risk factors and incident advanced liver disease in non-alcoholic fatty liver disease (NAFLD): A systematic review and meta-analysis of population-based observational studies. *PLoS Med*. 2020;17(4):e1003100. doi:10.1371/journal.pmed.1003100.
29. Sahlman P, Nissinen M, Puukka P, Jula A, Salomaa V, Männistö S, Lundqvist A, Liisa Valsta, Perola M, Färkkilä M, et al. Genetic and lifestyle risk factors for advanced liver disease among men and women. *J Gastroenterol Hepatol*. 2020;35(2):291-298.
30. Simon TG, Kim MN, Luo X, Yang W, Ma Y, Chong DQ, Fuchs CS, Meyerhardt JA, Corey KE, et al. Physical activity compared to adiposity and risk of liver-related mortality: Results from two prospective, nationwide cohorts. *J Hepatol*. 2020;72(6):1062-1069.
31. Stender S, Kozlitina J, Nordestgaard BG, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. *Nat Genet*. 2017;49(6):842-847.
32. Loomba R, Adams LA. Advances in non-invasive assessment of hepatic fibrosis. *Gut*. 2020;69(7):1343-1352.
33. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol*. 2018;69(1):182-236.

Table 1. Baseline characteristics by number of metabolic risk factors

Variable ¹	Metabolic risk factors			
	Three (n=21,059)	Two (n=220,462)	One (n=199,419)	Zero (n=45,888)
Age (SD), year	59.2 (9.3)	54.0 (10.6)	50.5 (10.6)	46.5 (8.3)
Female, %	65.5	62.9	56.9	49.8
Socioeconomic and lifestyle factors				
Urban region, %	62.2	43.7	44.3	33.1
≥9 years of education, %	20.3	21.0	21.7	17.9
Household income ≥35 000 RMB/year, %	20.3	19.2	17.5	13.5
Ever regular smoking, %				
Male	66.6	66.4	68.2	68.8
Female	3.3	2.9	3.0	1.9
Weekly drinking, %				
Male	30.7	35.6	32.3	32.1
Female	1.2	2.2	2.1	1.9
Total physical activity (SD), MET-h/day	14.7 (7.5)	16.3 (8.6)	22.5 (14.1)	41.9 (8.5)
Sedentary leisure time (SD), h/day	3.3 (1.7)	3.1 (1.6)	2.9 (1.5)	2.5 (1.3)
Blood pressure and anthropometry				
SBP (SD), mmHg	140.2 (22.5)	133.3 (21.5)	128.7 (20.4)	125.1 (18.3)
RPG (SD), mmol/L	12.7 (5.6)	6.0 (1.8)	5.6 (1.2)	5.5 (1.0)
BMI (SD), kg/m ²	25.7 (3.4)	24.9 (3.2)	22.5 (3.1)	21.5 (2.4)
Waist circumference (SD), cm	87.9 (8.8)	85.2 (8.4)	75.7 (8.4)	71.9 (6.1)
Hip circumference (SD), cm	93.1 (7.4)	92.4 (6.9)	89.6 (6.6)	87.7 (5.7)
Waist-to-hip ratio (SD)	0.94 (0.06)	0.92 (0.05)	0.84 (0.06)	0.81 (0.04)
Percent body fat (SD), %	31.9 (8.2)	30.5 (8.0)	25.6 (7.9)	23.6 (6.9)
Height (SD), cm	158.9 (8.4)	158.7 (8.5)	158.6 (8.0)	156.6 (7.7)
Prior disease history, %				
Diabetes	100.0	3.5	0.3	0.0
Coronary heart disease	5.9	3.3	2.4	1.1
Stroke or TIA	3.7	2.0	1.3	0.7
Hypertension	25.5	14.1	8.3	5.3
Family history of diabetes	12.6	5.1	4.4	4.5
Family history of cancer	14.7	14.0	13.7	13.7

Abbreviations: BMI=body mass index, MET=metabolic equivalent of task, RPG=random plasma glucose, SBP=systolic blood pressure, TIA=transient ischemic attack.

¹ Results were standardized by age, sex, and region (where appropriate). Values are means unless otherwise stated.

P-values of baseline characteristics between participants by numbers of metabolic risk factors: all <0.05.

Table 2. Associations between individual metabolic factors and risk of severe liver disease

Metabolic factor		Description	N cases/N total	HR (95% CI) ¹	HR (95% CI) ²
Central adiposity	No	WHR<0.90 (men), WHR<0.85 (women)	2059/285,220	1.00 (Ref)	1.00 (Ref)
	Yes	WHR≥0.90 (men), WHR≥0.85 (women)	1220/201,608	1.20 (1.11, 1.31)	1.18 (1.09, 1.28)
Physical activity	No	Total physical activity <33.2 MET-h/day	451/97,243	1.00 (Ref)	1.00 (Ref)
	Yes	Total physical activity ≥33.2 MET-h/day	2828/389,585	1.23 (1.11, 1.37)	1.22 (1.10, 1.36)
Hyperglycemia	No	Self-reported or screen-detected diabetes	330/28,715	1.00 (Ref)	1.00 (Ref)
	Yes	No diabetes	2949/458,113	1.49 (1.33, 1.68)	1.46 (1.30, 1.64)

Abbreviations: WHR=waist-to-hip ratio, MET=metabolic equivalent of task.

Cox proportional hazards regression models were used in the analysis.

¹ The model was adjusted for age at baseline, age squared, 10 regions, education, smoking, alcohol, and self-rated health.

² The model was additionally adjusted for the other two metabolic factors.

Table 3. Associations between combined metabolic risk factors and risk of severe liver disease

Metabolic factor	Severe liver disease		Liver cancer		Cirrhosis	
	N cases/ N total	HR (95% CI)	N cases/ N total	HR (95% CI)	N cases/ N total	HR (95% CI)
All participants						
3	253/21,059	1.00 (0.88, 1.14)	156/21,059	1.00 (0.85, 1.18)	132/21,059	1.00 (0.84, 1.20)
2	1635/220,462	0.69 (0.65, 0.73)	1008/220,462	0.71 (0.66, 0.76)	778/220,462	0.59 (0.55, 0.64)
1	1188/199,419	0.57 (0.53, 0.60)	747/199,419	0.61 (0.57, 0.67)	546/199,419	0.45 (0.41, 0.49)
0	203/45,888	0.48 (0.42, 0.56)	114/45,888	0.51 (0.42, 0.61)	102/45,888	0.37 (0.30, 0.45)
<i>per 1-point decrease</i>		<i>0.80 (0.75, 0.84)</i>		<i>0.82 (0.76, 0.88)</i>		<i>0.73 (0.67, 0.79)</i>
Population subset ¹						
3	40/3282	1.00 (0.72, 1.38)	26/3282	1.00 (0.67, 1.49)	17/3282	1.00 (0.61, 1.65)
2	227/32,294	0.64 (0.55, 0.73)	150/32,294	0.66 (0.56, 0.79)	98/32,294	0.61 (0.49, 0.76)
1	165/29,680	0.49 (0.41, 0.58)	103/29,680	0.54 (0.44, 0.67)	71/29,680	0.39 (0.31, 0.51)
0	28/6696	0.41 (0.28, 0.61)	15/6696	0.41 (0.24, 0.70)	15/6696	0.38 (0.22, 0.66)
<i>per 1-point decrease</i>		<i>0.75 (0.64, 0.87)</i>		<i>0.77 (0.64, 0.93)</i>		<i>0.69 (0.55, 0.86)</i>

Cox proportional hazards regression models were used in the analysis. The model was adjusted for age at baseline, age squared, 10 regions, education, smoking, alcohol, and self-rated health.

¹ This included a population-based sample of 75,736 participants randomly selected from the total CKB cohort and excluded 24,672 participants who had been selected for nested case-control studies. After excluding individuals with a prior history of cancer, cirrhosis or hepatitis, and positive HBsAg tests, 71,952 participants remained in the analysis.

Figure legends

Figure 1. Associations between combined metabolic risk factors and risk of severe liver disease by genetic factors

Boxes represent the hazard ratios (HRs) of SLD associated with combined metabolic risk factors in participants by strata of BMI GRS and rs738409, separately, with the size of the box inversely proportional to the variance of the logHR. Cox proportional hazards regression models were used in the analysis. The model was adjusted for age at baseline, age squared, 10 regions, education, smoking, alcohol, and self-rated health. *P*-values for interaction were obtained from the likelihood ratio test.

Figure 2. Associations between combined metabolic risk factors and liver biomarkers by genetic factors

Boxes represent the SD differences of liver biomarkers associated with combined metabolic risk factors in participants by strata of BMI GRS and rs738409, separately, with the size of the box inversely proportional to the variance of the SD difference. Linear regression models were used in the analysis. The model was adjusted for age at baseline, age squared, 10 regions, education, smoking, alcohol, and self-rated health. *P*-values for interaction were obtained from the likelihood ratio test. *P*-values for interaction: for FLI, BMI GRS 0.90 and rs738409 0.02; for GGT, BMI GRS 0.65 and rs738409 0.62.

Figure 1

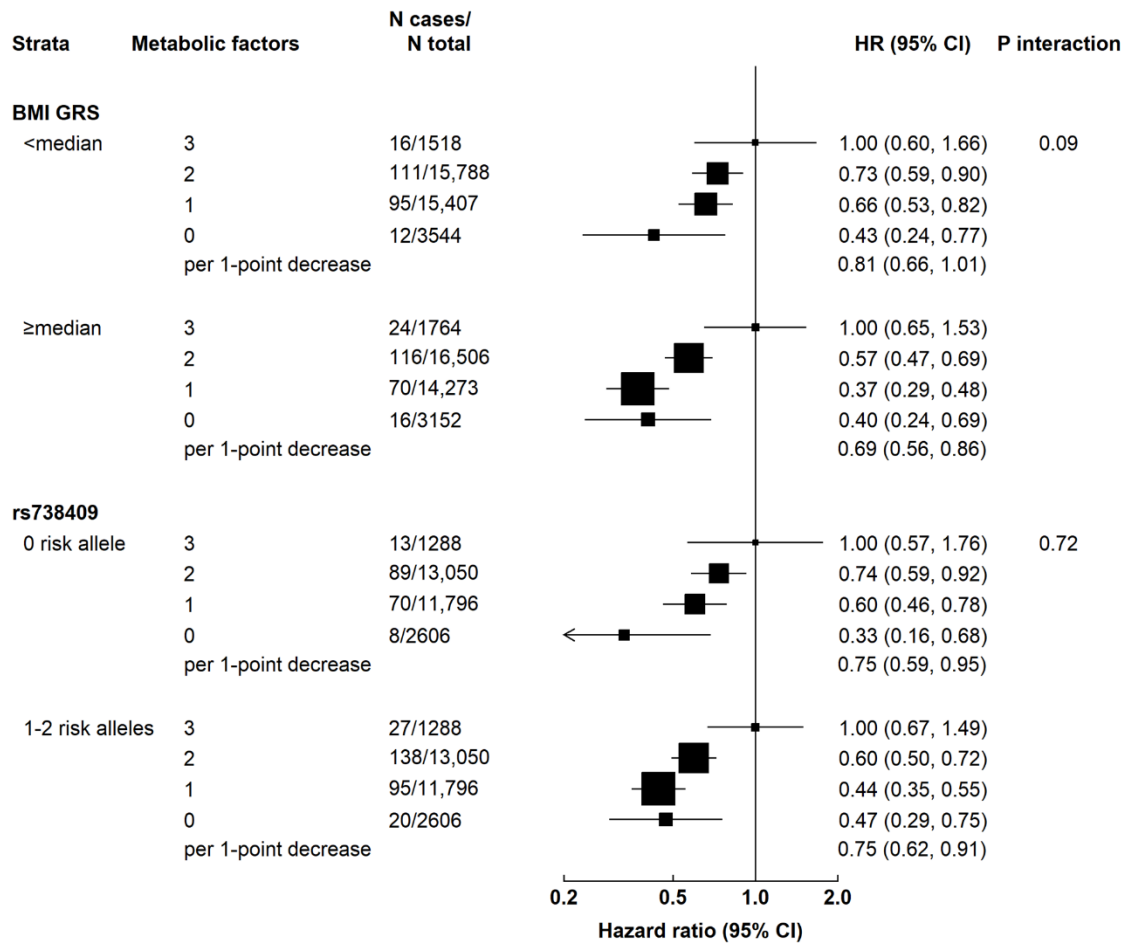


Figure 2

