Genetic risk, incident gastric cancer, and benefit from having a healthy lifestyle: a prospective cohort study

Guangfu Jin [*†], Jun Lv [*], Ming Yang [*], Mengyun Wang [*], Meng Zhu [*], Tianpei Wang [*], Caiwang Yan, Canqing Yu, Yanbing Ding, Gang Li, Chuanli Ren, Jing Ni, Ruoxin Zhang, Yu Guo, Zheng Bian, Yan Zheng, Nasha Zhang, Yue Jiang, Jiaping Chen, Yanong Wang, Dazhi Xu, Hong Zheng, Ling Yang, Yiping Chen, Robin Walters, Iona Y Millwood, Juncheng Dai, Hongxia Ma, Kexin Chen, Zhengming Chen, Zhibin Hu, Qingyi Wei [†], Hongbing Shen [†], Liming Li [†]

[*]Authors share co-first authorship.
[†] Senior authors who jointly directed the study contributed equally.

Department of Epidemiology, Center for Global Health, School of Public Health (Prof G Jin PhD, M Zhu PhD, T Wang MD, C Yan PhD, J Ni MD, Y Jiang MD, J Chen MS, J Dai PhD, Prof H Ma PhD, Prof Z Hu PhD, Prof H Shen PhD), and Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine and China International Cooperation Center for Environment and Human Health (Prof G Jin PhD, M Zhu PhD, C Yan PhD, Y Jiang MD, J Chen MS, J Dai PhD, Prof H Ma PhD, Prof Z Hu PhD, Prof H Shen PhD), Nanjing Medical University, Nanjing, China;

Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China (Prof J Lv PhD, C Yu PhD, Prof L Li MPH);

Shandong Provincial Key Laboratory of Radiation Oncology, Cancer Research Center, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China (Prof M Yang PhD, Y Zheng PhD, N Zhang PhD);

Cancer Institute, Fudan University Shanghai Cancer Center (M Wang PhD, R Zhang PhD, visiting Prof Q Wei PhD), and Department of Oncology, Shanghai Medical College (M Wang PhD, R Zhang PhD), Fudan University, Shanghai, China;

Department of Gastroenterology, the Affiliated Hospital of Yangzhou University, Yangzhou,

China (Prof Y Ding, PhD);

Department of General Surgery, Jiangsu Cancer Hospital, Jiangsu Institute of Cancer Research, The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, China (G Li PhD);

Department of Laboratory Medicine, Clinical Medical College of Yangzhou University, Yangzhou, China (C Ren PhD);

Chinese Academy of Medical Sciences, Beijing, China (Prof Y Guo MS, Z Bian MS);

Department of Gastric Surgery, Fudan University Shanghai Cancer Center, Shanghai, China, Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China (Y Wang MD, D Xu MD);

Department of Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China (H Zheng MD, Prof K Chen PhD);

Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK (L Yang PhD, Y Chen PhD, R Walters PhD, I Millwood PhD, Prof Z Chen PhD);

Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA (Prof Q Wei); Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, USA (Prof Q Wei).

[*]**Correspondence to:** Dr Hongbing Shen, Department of Epidemiology, School of Public Health, Nanjing Medical University, 101 Longmian Avenue, Nanjing 211166, China. E-mail: hbshen@njmu.edu.cn.

Summary

Background Genetic variants and lifestyle factors have been associated with risk of gastric cancer, but the extent to which an increased genetic risk can be offset by a healthy lifestyle remains unknown.

Methods We first constructed a PRS for gastric cancer from a meta-analysis of six genome-wide association studies of 10,254 gastric cancer cases and 10,914 cancer-free controls from Chinese populations. Then, we applied this PRS to 100,220 participants from the China Kadoorie Biobank cohort with more than ten years follow-up. The reduction in relative and absolute risk of incident gastric cancer associated with healthy lifestyles was assessed and stratified by genetic risk.

Results A significant increase in risk of incident gastric cancer was observed across the quintiles of PRS derived from 112 SNPs (P_{trend}<0·0001). Compared with participants at low genetic risk, those at intermediate and high genetic risk had a greater risk of gastric cancer, with a HR of 1·54 (95% CI, 1·22-1·94) and 2·08 (95% CI, 1·61-2·69), respectively; a similar increase in relative risk of incident gastric cancer was associated with lifestyle risk levels (P_{trend}<0·0001), with a 103% increased risk for an unfavorable lifestyle (HR, 2·03 [95% CI, 1·46-2·83]), compared with those with a favorable lifestyle. Participants at high genetic risk with a favorable lifestyle had a 47% lower gastric cancer risk (HR, 0·53 [95% CI, 0·29-0·99]), compared with those with an unfavorable lifestyle, and had an absolute risk reduction of 11·24‰ (95% CI, 6·24‰-15·61‰).

Interpretation Individuals at an increased risk of incident gastric cancer could be identified by a PRS. Individuals having a high genetic risk but a favorable lifestyle had a nearly 50% relative risk reduction.

Funding National Key R&D Program of China, National Natural Science Foundation of China, 333 High-Level Talents Cultivation Project of Jiangsu Province, and China Postdoctoral Science Foundation.

Research in context

Evidence before this study

We systematically searched PubMed and reviewed research articles published in English before Feb. 13th, 2020, with search terms including “polygenic risk score”, “healthy lifestyle” and “gastric cancer”. We found no studies that had constructed polygenic risk score (PRS) for gastric cancer to improve precision prevention of gastric cancer by performing a cohort study. When we searched “polygenic risk score” and “gastric cancer”, however, we found a study that had developed a PRS to determine risk of Barrett's esophagus or esophageal adenocarcinoma, which were more prevalent in European and American countries. However, whether such a PRS can identify individuals at high risk of gastric cancer remained unclear. Moreover, the effectiveness of PRS needed to be evaluated with additional independent large-scale cohorts.

Added value of this study

In the largest genetic association study of gastric cancer, we constructed a 112-SNP-based PRS for gastric cancer and evaluated its effectiveness with an independent large-scale prospective cohort in Chinese populations. This PRS successfully predicted incident rates of gastric cancer independent from lifestyle factors. Among participants at high genetic risk, a favorable lifestyle was associated with a lower gastric cancer risk, compared with an unfavorable lifestyle. These findings may facilitate individuals’ awareness of inherited susceptibility and participation in having a healthy lifestyle to prevent gastric cancer in high-risk populations.

Implications of all the available evidence

The 112-SNP-based PRS appears to be a practical and reliable genetic predictor for risk stratification of gastric cancer. Moreover, our results support for the benefit of adhering to a favorable lifestyle, especially in high genetic risk populations.

Introduction

Gastric cancer is one of the most common cancers and the third leading cause of cancer death in the world,[1] and more than 40% of new cases and deaths of gastric cancer occur in China.[2] Genetic predisposition, *Helicobacter pylori* (*H. pylori*) infection, and unhealthy lifestyles may independently and jointly contribute to the heavy burden of gastric cancer in Chinese populations.[3]

Unhealthy lifestyles, including tobacco smoking, alcohol drinking, consumption of salty or preserved foods and low intake of fresh vegetables or fruits, have been associated with an increased risk for gastric cancer. [4] Having a healthy lifestyle has been reported to reduce gastric cancer incidence by up to half in a Singapore Chinese population.[5] Much evidence has also shown that genetic risk may be attenuated by a favorable lifestyle for several common diseases, including coronary disease,[6] stroke,[7] dementia[8] and breast cancer[9]. However, the extent to which an increased genetic risk can be offset by a healthy lifestyle is unknown for gastric cancer.

A Nordic twins study has estimated that the heritability of gastric cancer is about 22%.[10] To date, a dozen of genetic variants have been found to be associated with gastric cancer risk, mostly in recent genome-wide association studies (GWASs).[11] Studies have shown that a polygenic risk score (PRS), by aggregating multiple common variants with a small effect into a risk model, efficiently predicts cancer risk and clinical outcomes.[12,13] However, it is critical to create a PRS to identify individuals at risk of gastric cancer for precision prevention.

Here, we constructed a PRS derived from gastric cancer GWASs of 10,254 cases and 10,914 controls and tested its utility and effectiveness in predicting risk of incident gastric cancer in a large cohort of 100,220 Chinese participants. We also assessed the extent to which having a healthy lifestyle may reduce risk of gastric cancer in Chinese populations, especially among those at a high genetic risk defined by the PRS.

Methods

Study design and participants

As shown in **Figure 1**, a two stage analysis was included in this study. In the first stage, we performed a meta-analysis of six independent GWAS datasets with a case-control study design, including two newly genotyped GWASs (the East-GWAS and the North-GWAS) and

four existing GWASs (the BJ-GWAS, the NJ-GWAS, the Onco-GWAS and the SX-GWAS). Then, we constructed five PRSs for gastric cancer with the summary statistics from the meta-analysis. In the second stage, we evaluated the effectiveness of PRSs and the impact of healthy lifestyle on gastric cancer risk in an independent prospective cohort of the China Kadoorie Biobank (CKB).

Cases and controls in each of the six GWAS datasets were all Han Chinese from the same geographical area (**appendix p 1**). For the East-GWAS and the North-GWAS, the cases were those who had histopathologically confirmed gastric cancer (aged more than 18 years) diagnosed at local hospitals, while cancer-free control subjects were recruited from individuals who participated in a community screening program for non-communicable diseases in Eastern (Shanghai and Jiangsu province) and Northern China (Shandong, Tianjin and Hebei provinces), respectively. Participants of the other four GWASs have been described in detail elsewhere.[11] All studies were approved by the relevant Institutional Review Boards (IRB), and all participants provided a written informed consent.

The CKB cohort is a prospective study whose study design, methods, and IRB approval have been described previously.[14, 15] In brief, a total of 512,714 adults aged 30-79 years were recruited from 10 geographically diverse (five urban and five rural) areas in China between June 2004 and July 2008 at baseline. Each eligible participant completed a written and IRB-approved informed consent form, an interviewer-administered electronic questionnaire, physical measurements, and a blood draw. The CKB cohort was approved by the Ethical Review Committee of the Chinese Center for Disease Control and Prevention (Beijing, China) and the Oxford Tropical Research Ethics Committee, University of Oxford (UK).

Procedures

We conducted quality control procedures on genotyping and imputation with the same protocol for all six GWAS datasets (**appendix p 2**). We performed association analyses separately for each GWAS dataset using the SNPTEST software (v.2.5).[16] We estimated odds ratios (ORs) and 95% confidence intervals (CIs) for genetic variants by using principal components of ancestry and demographic characteristics, including age, sex, smoking and alcohol drinking status (if available), as covariates in the multivariable logistic regression model.[17] The inverse variance-weighted fixed effects model was used in the meta-analysis as implemented by using the METAL software,[18] and the variants were excluded, if substantial

heterogeneity was present among studies ($P \leq 1\times10^{-4}$ for Cochran's Q test, or $I^2>0{\cdot}75$ as calculated by 100%×(Q-df)/Q).

We selected patients for inclusion in the calculation of PRS using a clustered random selection method from the CKB cohort, and their DNA samples were genotyped using a custom-designed Affymetrix Axiom array with approximately 700,000 markers. Individuals with cancers diagnosed at baseline were excluded from the analysis. Using summary statistics from the meta-analysis of GWASs, we derived a PRS by a pruning and thresholding approach with *P*-value and linkage disequilibrium (LD)-driven clumping procedure in LDpred.[19] In brief, summary statistics from the six GWAS meta-analysis were first coordinated with the CKB genotype data. We constructed five PRSs over a range of *P* value (5×10^{-4}, 5×10^{-5}, 5×10^{-6}, 5×10^{-7} and 5×10^{-8}) and r^2 (0·2) thresholds, which were generated by multiplying the dosage of the effect allele for each SNP by its respective weight. We incorporated genotype dosages for the uncertainty in the imputation and included the variants with sufficient imputation quality (INFO>0·5) for the scoring.

We adopted four healthy lifestyle factors according to the reported evidence for gastric cancer risk,[4] i.e., no current smoking, no alcohol drinking, low consumption of preserved foods and frequent intake of fresh fruits and vegetables. At baseline, participants of the CKB cohort attended survey clinics, where lifestyle factors were assessed by interviewers using laptop computer-based questionnaires. Participants of non-current smoking were defined as never smoker or former smokers who had quit smoking at least 15 years. No alcohol drinking was defined as no alcohol use in the past year and never drank in most weeks of the year. Low consumption of preserved foods was defined as eating preserved vegetables <4 days per week. The frequent intake of fresh fruits and vegetables was defined as eating vegetables very day and fruits ≥4 days per week or eating fruits very day and vegetables ≥4 days per week.

Outcomes of CKB gastric cancer cases were ascertained through ongoing electronic linkage with the established disease registries, the national health insurance claim database, and local death registries semi-annually. Participants who failed to be linked to the local health insurance database were actively followed annually for their vital status by the study staff, including hospital admission and residence of the study area. The staffs who were trained and blinded to baseline information coded the outcomes according to the 10th revision of the

International Classification of Diseases (ICD-10, http://www.who.int/classifications/icd/en/). Medical records were reviewed and necessary clinical information was collected by trained staff for diagnosis validation. The outcome in the final analysis was “C16-stomach cancer”.

Statistical analysis

Participants were considered at risk of gastric cancer from baseline and followed up until the date of first diagnosis, death, loss to follow-up, or the last date of hospital admission. Loss to follow-up referred to those could not be contacted after at least three times reasonable efforts within one year, or their new residence was out of the jurisdiction of the Regional Coordinating Center. By 31 December 2016, only about 1% participants were censored due to loss to follow-up. Multivariable logistic regression models were used to assess associations between the PRSs and individual lifestyle factors. A potential nonlinear relationship between the PRSs and gastric cancer risk was assessed by using the restricted cubic spline analysis.[20] Cox proportional hazards regression models were used to assess associations of genetic and lifestyle factors with gastric cancer incidence and to estimate hazards ratio (HR) and 95% confidence interval (CI) with adjustment for age, sex, residential area (urban or rural) and the first 10 principal components of ancestry[21-23]. The assumption of proportional hazards was assessed by testing the significance of Schoenfeld residuals. The PRSs were categorized into low (the bottom quintile), intermediate (quintiles 2-4) and high (the top quintile) genetic risk, as described previously.[6,8,24] We compared HRs for participants at high or intermediate genetic risk with those at low risk. Similarly, we compared favorable lifestyles (four healthy lifestyle factors) or intermediate lifestyles (2-3 factors) with unfavorable lifestyles (0 or 1 factor). We calculated absolute risk reduction as the difference in gastric cancer incidences among given groups, extrapolated the difference in 10-year event rates among given groups, and calculated the numbers of participates needed to adhere to favorable lifestyles to prevent one case of gastric cancer in ten years. The 95% CI of absolute risk reduction were generated by drawing 1000 bootstrap samples from the estimation dataset. To examine the robustness of the results, we conducted several sensitivity analyses: (1) defining genetic risk levels by different PRSs derived from a range of *P* values of the GWAS meta-analysis; (2) reclassifying genetic risk levels based on quartiles (low: the bottom quartile; intermediate: quartiles 2-3; and high: the top quartile) or tertiles (the lowest tertiles, the mid tertiles, and the highest tertiles) of PRSs; (3) reclassifying lifestyles when the frequent intake of fresh fruits and vegetables was defined as eating vegetables and fruits very day; and (4) excluding incident cases of gastric cancer occurring during the first year of follow-up. A two-sided p<0·05 was considered statistically

significant. All analyses were performed with R software, version 3.1 (R Project for Statistical Computing).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

The meta-analysis included 10,254 cases and 10,914 controls from six GWAS datasets (**appendix p 3**) with about seven-million genetic variants at a minor allele frequency no less than 0.5% (**appendix p 18-22**) after imputation. We found that 764 variants within six regions (per 1000 kb apart) were significantly associated with gastric cancer risk (P=5×10^{-8}) (**appendix p 23**). Because aggregating more common variants into a single genetic model may improve the power of predicting genetic risk,[25, 26] we constructed five PRSs in a range of significance thresholds (5×10^{-4}, 5×10^{-5}, 5×10^{-6}, 5×10^{-7} and 5×10^{-8}), including 12, 18, 38, 112, and 539 independent SNPs (**appendix p 4**), respectively, on the basis of their LD patterns.

A total of 100,641 individuals were selected and genotyped from the CKB cohort, and 421 individuals with cancers diagnosed at baseline were excluded, resulting in 100,220 participants included in the final analysis. Among these participants, 692 incident gastric cancer cases were ascertained during a median follow-up of 10·4 years (IQR 9·34-11·30 years) (**Table 1**). We calculated the aforementioned five PRSs for each participant by using summary statistics from the GWAS meta-analysis. All five PRSs had an approximate normal distribution (**appendix p 24**) and were significantly associated with gastric cancer risk (**appendix p 4**). Because the PRS derived from 112 SNPs (**appendix p 5**) with a P value less than 5×10^{-5} showed the strongest association (P=7·56×10^{-10}) and obviously different distributions between cases and controls in GWASs (**appendix p 25**), we used this PRS to define genetic risk for subsequent analyses. The other PRSs were used for sensitivity analyses to evaluate the stability of genetic predictors for risk of gastric cancer in the CKB cohort.

The PRS derived from 112 SNPs was not associated with any of the lifestyle factors or other variables collected at baseline of the CKB cohort (**appendix p 8**). The participants with incident gastric cancer tended to have a higher PRS (genetic risk) than those unaffected

(**appendix p 26**). There was an obvious linear relationship for the positive association between PRS and gastric cancer risk, with an HR of 1·27 (95%CI, 1·17-1·37, $P_{overall}$<0·0001 and $P_{nonlinear}$=0·777) per SD of PRS increase (**appendix p 26**). We observed a significantly increasing risk for incident gastric cancer across the quintiles of genetic risk (P_{trend}<0·0001; **Figure 2A**), which did not change with additional adjustment for lifestyle factors (P_{trend} <0·0001; **appendix p 9**). Compared with participants at low genetic risk, participants at intermediate and high genetic risk had a significantly higher risk of gastric cancer, with HRs of 1·54 (95%CI, 1·22-1·94) and 2·08 (95% CI, 1·61-2·69), respectively (**Figure 2B**). For example, the absolute cumulative risk of a 60-year-old male developing gastric cancer over the next 10 years is 7.5‰, 11.5‰, and 15.6‰ for low, intermediate, and high genetic risk individuals, respectively, on average. These results also did not change after additional adjustment for lifestyle factors (**appendix p 10**). Similarly, genetic risk reclassified by quartiles or tertiles of the PRS or categorized by the other PRSs also showed an association with gastric cancer risk independent of lifestyle factors (**appendix p 11-12**). Furthermore, the association of genetic risk with incident gastric cancer was unchanged, when we excluded incident cases during the first year of follow-up (**appendix p 13**).

In the CKB cohort, most participants had two (25,335 [25·28%] of 100,220 participants) or three (42,935 [42·84%] of 100,220 participants) of four healthy lifestyles (**Table 1** and **appendix p 14**). There was a significant increase in the strength of association with gastric cancer risk as the number of healthy lifestyles increased (P_{trend}<0·0001; **Figure 2C**). We then divided participants by lifestyle into three levels: favorable (13,576 [13·55%] of 100,220 participants), intermediate (68,270 [68·12%] of 100,220 participants) and unfavorable (18,374 [18·33%] of 100,220 participants). A significantly increasing gastric cancer risk was also observed across healthy lifestyle levels (P_{trend}<0·0001, **Figure 2D**). The relative risk of incident gastric cancer was 103% greater among participants with an unfavorable lifestyle (HR, 2·03 [95%CI, 1·46-2·83]) than among those with a favorable lifestyle, and additional adjustment for genetic risk resulted in an HR of 2·02 (95%CI, 1·45-2·82) (**appendix p 15**). Similar results were observed when frequent intake of fresh fruits and vegetables was redefined (**appendix p 16**).

We also observed a joint effect of genetic and lifestyle factors on risk of incident gastric cancer in a dose-response manner; that is, the overall risk of gastric cancer increased as both genetic risk and unhealthy lifestyle-related risk increased (**Figure 3**). Specifically,

participants with a high genetic risk and an unfavorable lifestyle had the highest risk of incident gastric cancer (HR, 5·14 [95%CI, 2·04-12·93]) than participants with a low genetic risk and a favorable lifestyle; however, there was no significant interaction between genetic risk and lifestyle factors ($P = 0·451$).

In further stratification analyses by genetic risk using unfavorable lifestyles as the reference group, we found that there was a significant benefit in the relative risk reduction of gastric cancer from having a favorable lifestyle across genetic risk groups (**Table 2** and **appendix p 27**). Specifically, among participants at high genetic risk, those with a favorable lifestyle had a 47% reduced gastric cancer risk (HR, 0·53 [95%CI, 0·29-0·99]), compared with those with an unfavorable lifestyle (**Table 2**). There was a 10-year absolute risk reduction of 11·24‰ (95%CI, 6·24‰-15·61‰) or 8.18‰ (95%CI, 4·50‰-11·94‰) associated with a favorable lifestyle in participants at high or low genetic risk, respectively (**Table 2**). Correspondingly, the number of participants needed to adhere to favorable lifestyles to prevent one incident gastric cancer case in 10 years was 89 and 122 for high and low genetic risk levels, respectively (**Table 2**). We repeated the analyses by reclassifying genetic risk levels according to quartiles or tertiles of the PRS and observed a similar risk-reduction benefit from having a favorable lifestyle, compared with an unfavorable lifestyle (**appendix p 17**).

Discussion

To date, PRSs derived from genetic variants identified in GWASs, mostly conducted in European populations, have reportedly led to significant disparities in prediction accuracy across diverse populations.[13, 27] Therefore, those findings from European populations may not be applied to precision prevention in other ethnic populations. In the present study, we integrated the largest GWAS of gastric cancer in Chinese populations to maximize the statistical power and provided an example of using the PRS to predict and assess cancer risk in Chinese populations. Because the incidence of gastric cancer is high in East Asian countries but rare in West developed countries, gastric cancer has been inadequately studied in populations of European ancestry.[28] Therefore, our findings provided the first evidence, to the best of our knowledge, for PRS to be used in predicting genetic risk of gastric cancer in Chinese populations.

A key public health need is to identify individuals at high risk of gastric cancer for primary (for example, *H. pylori* eradication) and secondary preventions (for example, enhanced

endoscopic screening). Recently, a risk prediction score summarizing seven predictors (age, sex, serum pepsinogen I/II ratio, gastrin-17 level, *H. pylori* infection, and pickled food and fried food consumption) was used to identify individuals at high risk of gastric cancer for endoscopic screening in a Chinese population.[29] Our results further indicated that a PRS derived from a number of risk-predictive SNPs identified in GWASs could predict individuals at-risk of incident gastric cancer, independent of lifestyle factors. Our findings are consistent with those of previous studies of using PRSs as robust risk predictors for other cancers[9, 26] or complex diseases.[6-8] These studies, including ours, collectively support that genetic risk of many common diseases, including cancer, may help facilitate individuals' awareness of their inherited susceptibility to common diseases and their participation in precision prevention. In the East Asian countries where the incidence of gastric cancer is remarkably high, the personalized prevention measures may be applied to individuals at high-risk, such as test for and eradication of *H. pylori* or more frequent proactive endoscopy screening.

Healthy lifestyles have been inversely associated with risk of gastric cancer,[5] and the present study further showed that healthy lifestyles could reduce gastric cancer risk across each level of genetic risk as defined by the PRS, and individuals with high genetic risk would also benefit from adopting healthy lifestyles. Therefore, our results support the notion that public efforts in emphasizing healthy lifestyles for everyone will lead to a reduction of gastric cancer risk.

Nevertheless, additional studies are needed to address several concerns in perfecting the existing prediction model. First, because we used PRSs derived from many SNPs identified in GWASs to assess genetic risk of gastric cancer only in Chinese populations, the results may not be generalizable to other ethnic populations. Second, although the eradication of *H. pylori* may reduce risk of gastric cancer,[30] some debate continues about what is the optimal population for *H. pylori* eradication.[31] Therefore, an evaluation of the benefit from *H. pylori* eradication in populations at different levels of genetic risk could be extremely useful for an individualized primary prevention of gastric cancer. Third, an endoscopic screening has been recommended to detect gastric cancer at earlier stages that have a more favorable prognosis.[32, 33] However, it is unclear whether individuals at a high genetic risk could have a maximized benefit from screening than those with a low genetic risk, which could finally improve secondary prevention of gastric cancer.

The present study has several limitations. First, the status of *H. pylori* infection at baseline was unavailable in the CKB cohort, which precluded us from evaluating the utility and effectiveness of PRS stratified by *H. pylori* status. Second, lifestyle factors were self-reported with only the frequency of diet recorded, which might result in misclassification of lifestyle risk levels. Third, lifestyle factors were assessed at baseline, and changes in behaviors during the follow-up might have an effect on risk estimates. Fourth, we only evaluated overall gastric cancer risk, but genetic and lifestyle factors might differ by tumor location (cardia or non-cardia) or subtype (diffuse or intestinal) of gastric cancer that were unavailable in the present study. At last, imputation was used for missing data on genotypes among different chips, which might lead to partial deviation of our estimates, even though only variants with high quality were kept.

In conclusion, our results of PRSs derived from the largest GWAS of gastric cancer in Chinese populations and subsequent validation in a large Chinese cohort study demonstrated that genetic and lifestyle factors were independently associated with risk of incident gastric cancer. Among those at high genetic risk, having a favorable lifestyle led to a nearly 50% reduction in the relative risk of gastric cancer, compared with those with an unfavorable lifestyle, highlighting that genetically susceptible individuals, once identified, could partially offset their risk of gastric cancer by being advised to adhere to healthy lifestyles.

Authors' Contributions

GJ, JL, QW, HS, and LL contributed to the study design and sample collection, and supervised the whole project. GJ, MZ, TW, and HS contributed to the data interpretation, data analysis, and writing of the manuscript. MY, MW, Caiwang Yan, Canqing Yu, YD, GL, CR, JN, RZ, YG, ZB, YZ, NZ, YJ, JC, YW, DX, HZ, LY, YC, RW, IYM, JD, HM, ZW, KC, ZC, and ZH contributed to the study design, sample collection, experiment or data interpretation of the present analysis. All of the authors reviewed or revised the manuscript.

Declaration of interests

We declare no competing interests.

Acknowledgments

This study was funded by National Key R&D Program of China (2016YFC1302703); National Natural Science Foundation of China (81872702 and 81521004); 333 High-Level Talents Cultivation Project of Jiangsu Province (BRA2018057); Project funded by China Postdoctoral Science Foundation (2019TQ0157, 2018M640466). None of the external funders or sponsors had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. We thank all the study participants and research staff for their contributions and commitment to the present study. The chief acknowledgment is to the participants, the project staff, and the China National Centre for Disease Control and Prevention (CDC) and its regional offices for assisting with the fieldwork.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424. doi: 10.3322/caac.21492
2. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115-132. doi: 10.3322/caac.21338
3. Chen W, Xia C, Zheng R, et al. Disparities by province, age, and sex in site-specific cancer burden attributable to 23 potentially modifiable risk factors in China: a comparative risk assessment. *Lancet Glob Health*. 2019;7(2):e257-e269. doi: 10.1016/S2214-109X(18)30488-1
4. World Cancer Research Fund International. Diet, nutrition, physical activity and stomach cancer. https://www.wcrf.org/sites/default/files/Stomach-Cancer-2016-Report.pdf Accessed December 24, 2019.
5. Wang Z, Koh WP, Jin A, Wang R, Yuan JM. Composite protective lifestyle factors and risk of developing gastric adenocarcinoma: the Singapore Chinese Health Study. *Br J Cancer*. 2017;116(5):679-687. doi: 10.1038/bjc.2017.7
6. Khera AV, Emdin CA, Drake I, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med*. 2016;375(24):2349-2358. doi: 10.1056/NEJMoa1605086
7. Rutten-Jacobs LC, Larsson SC, Malik R, et al. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. *BMJ*. 2018;363:k4168. doi: 10.1136/bmj.k4168
8. Lourida I, Hannon E, Littlejohns TJ, et al. Association of Lifestyle and Genetic Risk with Incidence of Dementia [published online July 15,2019]. *JAMA*. doi: 10.1001/jama.2019.9879
9. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol*. 2016;2(10):1295-1302. doi: 10.1001/jamaoncol.2016.1025
10. Mucci LA, Hjelmborg JB, Harris JR, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*. 2016;315(1):68-76. doi: 10.1001/jama.2015.17703
11. Yan C, Zhu M, Ding Y, et al. Meta-analysis of genome-wide association studies and functional assays decipher susceptibility genes for gastric cancer in Chinese populations

[published online Aug 5,2019]. *Gut*. doi: 10.1136/gutjnl-2019-318760

12. Sugrue LP, Desikan RS. What Are Polygenic Scores and Why Are They Important? *JAMA*. 2019;321(18):1820-1821. doi: 10.1001/jama.2019.3893
13. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584-591. doi: 10.1038/s41588-019-0379-x
14. Chen Z, Lee L, Chen J, et al. Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC). *Int J Epidemiol*. 2005;34(6):1243-1249. doi: 10.1093/ije/dyi174
15. Chen Z, Chen J, Collins R, et al. 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. doi: 10.1093/ije/dyr120
16. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39(7):906-913. doi: 10.1038/ng2088
17. Peduzzi P, Concato J. Kemper E, Holfors TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996; 49(12):1373-1379. doi: 10.1016/s0895-4356(96)00236-3
18. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi: 10.1093/bioinformatics/btq340
19. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet*. 2015;97(4):576-592. doi: 10.1016/j.ajhg.2015.09.001
20. Durrleman S, Simon R. Flexible regression models with cubic splines. *Stat Med*. 1989;8(5):551-561. doi: 10.1002/sim.4780080504
21. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38(8):904-909. doi: 10.1038/ng1847
22. Concato J, Peduzzi P, Holford TH, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals and general strategy. *J Clin Epidemiol.* 1995; 48(12):1495-1501. doi: 10.1016/0895-4356(95)00510-2
23. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards analysis. II. Accuracy and precision of regression

estimates. *J Clin Epidemiol.* 1995; 48(12):1503-1510. doi: 10.1016/0895-4356(95)00048-8

24. Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet*. 2015; 385(9984):2264-2271. doi: 10.1016/S0140-6736(14)61730-X
25. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*. 2016;17(7):392-406. doi: 10.1038/nrg.2016.27
26. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. doi: 10.1038/s41588-018-0183-z
27. Gurdasani D, Barroso I, Zeggini E, et al. Genomics of disease risk in globally diverse populations. *Nat Rev Genet*. 2019;20(9):520-535. doi: 10.1038/s41576-019-0144-0
28. Mills JC, Samuelson LC. Past Questions and Current Understanding About Gastric Cancer. *Gastroenterology*. 2018;155(4):939-944. doi: 10.1053/j.gastro.2018.06.044
29. Cai Q, Zhu C, Yuan Y, et al. Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study. *Gut*. 2019;68(9):1576-1587. doi: 10.1136/gutjnl-2018-317556
30. Lee YC, Chiang TH, Chou CK, et al. Association Between Helicobacter pylori Eradication and Gastric Cancer Incidence: A Systematic Review and Meta-analysis. *Gastroenterology*. 2016;150(5):1113-1124.e5. doi: 10.1053/j.gastro.2016.01.028
31. Sugano K, Tack J, Kuipers EJ, et al. Kyoto global consensus report on Helicobacter pylori gastritis. *Gut*. 2015;64(9):1353-1367. doi: 10.1136/gutjnl-2015-309252
32. Hamashima C, Systematic Review Group and Guideline Development Group for Gastric Cancer Screening Guidelines. Update version of the Japanese Guidelines for Gastric Cancer Screening. *Jpn J Clin Oncol*. 2018;48(7):673-683. doi: 10.1093/jjco/hyy077
33. Kim Y, Jun JK, Choi KS, Lee HY, Park EC. Overview of the National Cancer screening programme and the cancer screening status in Korea. *Asian Pac J Cancer Prev.* 2011;12(3):725-730

Figure Legends

Figure 1. Study design and workflow

Figure 2. The relationships of genetic and lifestyle factors with incident gastric cancer in the China Kadoorie Biobank (CKB) cohort. (A) Participants in the CKB cohort were divided into five equal groups according to the PRS, and the hazard ratios (HRs) of each group were compared with those in the bottom quintile of the polygenic risk score (PRS). Error bars are 95% CIs. (B) Standardized rates of gastric cancer events in low (bottom quintile), intermediate (quintiles 2-4), and high (top quintile) genetic risk groups in the CKB cohort. (C) Participants in the CKB cohort were divided into five groups according to the numbers of healthy lifestyle factors and the hazards ratios (HRs) of each group were compared with those having 4 healthy lifestyle factors. Error bars are 95% CIs. (D) Standardized rates of gastric cancer events in favorable (4 healthy lifestyle factors), intermediate (2-3 healthy lifestyle factors), and unfavorable (0 or 1 healthy lifestyle factor) lifestyle groups in the CKB cohort. HRs and 95% CIs were estimated using Cox proportional-hazard models with adjustment for age, sex, residential area, and the first 10 principal components of ancestry. Shaded areas are 95% CIs.

Figure 3. Risk of incident gastric cancer according to genetic and lifestyle categories. The hazard ratios were estimated using Cox proportional-hazard models with adjustment for age at risk, sex, residential area, and the top 10 principal components of ancestry.