# Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Du H, Li L, Bennett D, et al. Fresh fruit consumption and major cardiovascular disease in China. N Engl J Med 2016;374:1332-43. DOI: 10.1056/NEJMoa1501451

**Table of Contents** (Click on an underlined item to jump to it)

Page 2. List of study committees and investigators.

Page 3. Calculation of the usual amount of fresh fruit consumption.

Page 5. Table S1. Baseline and usual portions per month for each baseline category of fruit consumption based on 17,488 participants who attended the first resurvey in 2008

Page 6. Detailed statistical analyses.

Page 8. Table S2. Repeatability of fresh fruit consumption assessed by simple qualitative food frequency questionnaire on two occasions.

Page 9. Table S3. Adjusted hazard ratios (HRs) of cardiovascular mortality and incidence of major cardiovascular diseases for daily versus never/rarely fresh fruit consumption, with different adjustment for other dietary factors.

Page 10. Table S4. Agreement between consumption frequency and quantitative categories integrating daily portion of consumption among those participated in the second resurvey.

Page 11. Figure S1. Percentage of participants who reported consuming fresh fruit on most days by age

Page 12. Figure S2. Percentage of participants who reported consuming fresh fruit on most days by age and survey area.

Page 13. Figure S3. Percentage of participants consuming fresh fruit on most days by area.

Page 14. Figure S4. Adjusted hazard ratios (HRs) of other cardiovascular diseases by usual habitual levels of fresh fruit consumption.

Page 15. Figure S5. Adjusted hazard ratios (HRs) of major cardiovascular diseases per 1 daily portion of fresh fruit consumption by baseline characteristics.

Page 16. Figure S6. Adjusted hazard ratios (HRs) of other cardiovascular diseases per 1 daily portion of fresh fruit consumption by baseline characteristics.

Page 17. Figure S7. Adjusted hazard ratios (HRs) of major cardiovascular events per 1 daily portion of fresh fruit consumption by region.

Page 18. Figure S8. Adjusted hazard ratios of major cardiovascular diseases per 1 daily portion of fresh fruit consumption by season of enrolment.

Page 19. Figure S9. Adjusted hazard ratios (HRs) of major cardiovascular diseases per 1 daily portion of fresh fruit consumption with different degree of adjustment for potential mediators.

Page 20. Supplementary References

**List of Study Committees and Investigators in the China Kadoorie Biobank Collaborative Group**

**International Steering Committee:** Junshi Chen, Zhengming Chen (PI), Rory Collins, Liming Li (PI), Richard Peto. **International Co-ordinating Centre, Oxford:** Daniel Avery, Derrick Bennett, Yumei Chang, Yiping Chen, Zhengming Chen, Robert Clarke, Huaidong Du, Xuejuan Fan, Haiyan Gao, Simon Gilbert, Michael Holmes, Andri Iona, Rene Kerosi, Ling Kong, Om Kurmi, Garry Lancaster, Sarah Lewington, John McDonnell, Winnie Mei, Iona Millwood, Qunhua Nie, Jayakrishnan Radhakrishnan, Paul Ryder, Sam Sansome, Dan Schmidt, Paul Sherliker, Rajani Sohoni, Robin Walters, Jenny Wang, Lin Wang, Alex Williams, Ling Yang, Xiaoming Yang. **National Co-ordinating Centre, Beijing:** Zheng Bian, Ge Chen, Lei Guo, Yu Guo, Bingyang Han, Can Hou, Peng Liu, Jun Lv, Pei Pei, Shuzhen Qu, Yunlong Tan, Canqing Yu, Huiyan Zhou. **10 Regional Co-ordinating Centres: Qingdao** Qingdao CDC: Zengchang Pang, Shutao Pang, Shaojie Wang, Yongmei Liu, Ranran Du, Yajing Zang, Liang Cheng. Licang CDC: Silu Lv, Junzheng Wang, Wei Hou. **Heilongjiang** Provincial CDC: Jiyuan Yin, Shumei Liu, Zhigang Pang, Xue Zhou, Huijun Wang. Nangang CDC: Liqiu Yang, Bo Yu, Yanjie Li, Jing Qi, Huaiyi Mu, Qin'ai Xu, Meiling Dou. **Hainan** Provincial CDC: Jianwei Du, Shanqing Wang, Ximin Hu, Hongmei Wang, Jinyan Chen, Yan Fu, Zhenwang Fu, Xiaohuan Wang, Hua Dong. Meilan CDC: Min Weng, Xiangyang Zheng, Yijun Li, Huimei Li. **Jiangsu** Provincial CDC: Ming Wu, Jinyi Zhou, Ran Tao, Jie Yang. Suzhou CDC: Jie Shen, Yihe Hu, Yan Lu, Yan Gao, Liangcai Ma, Aiyu Tang, Shuo Zhang, Jianrong Jin. **Guangxi** Provincial CDC: Zhenzhu Tang, Naying Chen, Ying Huang. Liuzhou CDC: Mingqiang Li, Jinhuai Meng, Rong Pan, Qilian Jiang, Jingxin Qin, Weiyuan Zhang, Yun Liu, Liuping Wei, Liyuan Zhou, Ningyu Chen, Jun Yang, Hairong Guan. **Sichuan** Provincial CDC: Xianping Wu, Ningmei Zhang, Xiaofang Chen, Xuefeng Tang. Pengzhou CDC: Guojin Luo, Jianguo Li, Xiaofang Chen, Jian Wang, Jiaqiu Liu, Qiang Sun. **Gansu** Provincial CDC: Pengfei Ge, Xiaolan Ren, Caixia Dong. Maiji CDC: Hui Zhang, Enke Mao, Xiaoping Wang, Tao Wang. **Henan** Provincial CDC: Guohua Liu, Baoyu Zhu, Gang Zhou, Shixian Feng, Liang Chang, Lei Fan. Huixian CDC: Yulian Gao, Tianyou He, Li Jiang, Huarong Sun, Pan He, Chen Hu, Qiannan Lv, Xukui Zhang. **Zhejiang** Provincial CDC: Min Yu, Ruying Hu, Le Fang, Hao Wang. Tongxiang CDC: Yijian Qian, Chunmei Wang, Kaixue Xie, Lingli Chen, Yaxing Pan, Dongxia Pan. **Hunan** Provincial CDC: Yuelong Huang, Biyun Chen, Donghui Jin, Huilin Liu, Zhongxi Fu, Qiaohua Xu. Liuyang CDC: Xin Xu, Youping Xiong, Weifang Jia, Xianzhi Li, Libo Zhang, Zhe Qiu.

**Calculation of the usual amount of fresh fruit consumption**

The group mean level of usual amount of fresh fruit consumption (in portions per month), was calculated for each baseline consumption group, using the following steps:

1. The mean daily portions for each of the five fresh fruit consumption groups namely 'never/rarely', 'monthly', '1-3 days per week', '4-6 days per week' or 'daily' were computed using data collected during the 2nd resurvey. These were found to be 0 for the 'never/rarely' group, 1.37 for the 'monthly' group, 1.43 for the '1-3 days/week' group, 1.40 for the '4-6 days/week' group, and 1.68 for the 'daily' group.
2. The mean portions per month of fresh fruit consumption were derived using the mean daily portions multiplied by the mean number of days per month. The mean number of days assigned to each group was 0 for the 'never/rarely' group, 2.5 days for the 'monthly' group, 8.6 days for the '1-3 days/week' group, 21.5 days for the '4-6 days/week' group, and 30 days for the 'daily' group. Using these values lead to following estimates of mean portions per month of fruit consumption of 0, 3.4, 12.3, 30.1, and 50.4, respectively, for 'never/rarely', 'monthly', '1-3 days per week', '4-6 days per week' or 'daily' fresh fruit consumption groups.
3. Assuming that the daily portions of fresh fruit consumption did not vary much from the baseline to the 2nd resurvey (i.e. people may change their frequency of fruit consumption but the daily amount of fresh fruit consumption remains roughly the same), the mean portions per month derived in the Step 2 was used as a proxy for the '***baseline amount of fresh fruit consumption***'.
4. The information from the 1st resurvey (Table S1) was used to assess the variation in fresh fruit consumption from the baseline. The 1st resurvey was conducted at roughly the mid-point of the follow-up period. The average '***usual amount of fresh fruit consumption***' for each baseline group ('never/rarely', 'monthly', '1-3 days per week', '4-6 days per week' or 'daily') was then calculated by multiplying the '***baseline amount of fresh fruit consumption***' for each group (calculated in the Step 2) by the percentage of participants in each of the consumption frequency group at the 1st resurvey (Appendix S2 Table on page 5). For example, for the baseline 'never/rarely' group, at the 1st resurvey 17.76% remained at the 'never/rarely' group , 56.27% switched to the 'monthly' group, 17.59% switched to the '1-3 days per week' group, 4.15% switched to the '4-6 days per week' group, and 4.23% switched to the 'daily' group. The **'*usual amount of fresh fruit consumption*'** for the baseline 'never/rarely' group was therefore

equal to 17.76%*0 + 56.27%*3.4 + 17.59%*12.3 + 4.15%*30.1 + 4.23%*50.4 = 7.47 portions per month.

5. The ***'usual amount of fresh fruit consumption'*** for the other baseline categories of fresh fruit consumption (i.e. 'monthly', '1-3 days per week', '4-6 days per week' or 'daily') were calculated using the same approach, leading to 9.81, 16.63, 23.62 36.33 portions per month respectively. This ***'usual amount fresh fruit consumption'*** was then assigned to each individual participant in order to compute an effect size per 1 portion/day of fresh fruit consumption.
6. The average 'usual amount fresh fruit consumption' was also used to plot against the HRs computed via Cox regression analyses with baseline fruit consumption as a categorical variable.

**Table S1. Baseline and usual portions per month for each baseline category of fruit consumption based on 17,488 participants who attended the first resurvey in 2008**

| Baseline | | 1st resurvey (mean 2.6 years later) **F** | | | | | Consumption days per month | Mean daily portion* | Baseline consumption (portions/month) **B** | Usual consumption (portions/month)† **U** |
|---|---|---|---|---|---|---|---|---|---|---|
| n | | Never/rarely 1 | Monthly 2 | 1-3 days/wk 3 | 4-6 days/wk 4 | Daily 5 | | | | |
| 1 | Never/rarely | 17.76 % (214) | 56.27% (678) | 17.59% (212) | 4.15% (50) | 4.23% (51) | 0 | - | $B_1 = 0$ | $U_1 = 7.47$ |
| 2 | Monthly | 7.10% (433) | 51.94% (3169) | 29.44% (1796) | 6.92% (422) | 4.61% (281) | 2.5 | 1.37 | $B_2 = 3.4$ (2.5×1.37) | $U_2 = 9.81$ |
| 3 | 1-3 days/wk | 3.90% (216) | 31.88% (1766) | 37.86% (2097) | 11.84% (656) | 14.52% (804) | 8.6 | 1.43 | $B_3 = 12.3$ (8.6×1.43) | $U_3 = 16.63$ |
| 4 | 4-6 days/wk | 1.53% (25) | 20.09% (329) | 35.47% (581) | 15.08% (247) | 27.84% (456) | 21.5 | 1.40 | $B_4 = 30.1$ (21.5×1.40) | $U_4 = 23.62$ |
| 5 | Daily | 1.30% (39) | 7.25% (218) | 19.00% (571) | 13.70% (411) | 58.77% (1766) | 30 | 1.68 | $B_5 = 50.4$ (30×1.68) | $U_5 = 36.33$ |

*The mean daily portion number came from the 2nd resurvey data, used as a proxy of baseline mean daily portion.

†Usual intake amount for each group was estimated by taking into account changes in consumption frequency between baseline and 1st resurvey using this formula $Un = \sum_{i=1}^{5}(Fni \times Bi)$; F is the percentage in each cell, B is the baseline proportion per month for each baseline category, U is the usual proportion per month for each baseline category.

**Detailed statistical analysis**

For the present study we included 451,665 participants who were free of prior cardiovascular disease (n=23,132) or anti-hypertensive treatments (n=48,174) at baseline. Weighted kappa and Spearman correlation coefficients were calculated to assess the reliability of fruit consumption categorization. Baseline characteristics were categorized by levels of fruit consumption using means and standard deviations or percentages, with adjustment for age, sex, and region, where appropriate, using either multiple linear regression (for continuous outcomes) or logistic regression (for binary outcomes). The marginal mean levels and 95% confidence intervals (CIs) of body mass index (BMI), waist circumference, blood pressure, and blood glucose by fruit consumption were estimated separately for men and women using multiple linear regression models, adjusting for age (10 categories), area (10 regions), education (4 categories), annual household income (4 categories), smoking (4 categories), alcohol intake (4 categories), physical activity (continuous variable), consumption of meat (3 categories), dairy products (3 categories) and preserved vegetables (5 categories), and survey season (4 categories).

Cox regression was used to calculate hazard ratios (HRs) and 95% CIs of fruit consumption in relation to disease risks, adjusting for the above-mentioned potential confounders and stratified by age-at-risk (in 5-year intervals), sex, and region. The proportionality assumption was checked by comparing the HRs for the first and second half of the follow-up period. For analyses involving more than two exposure categories, the floating absolute risk method was used, which provides variances of log risk for each category (including the reference group) to facilitate comparisons between different exposure groups.[1] For comparison, the conventional (un-floated) analyses were also performed. The HR of cardiovascular death for non-daily versus daily fruit consumption was calculated in order to estimate the population-attributable fraction using the formula: *Pe(HR-1)/[Pe(HR-1)+1],*[2] where Pe denotes the prevalence of non-daily fresh fruit consumption in the Chinese population.

To further quantify the association between amount of fruit intake and disease risks and to account for regression dilution bias,[3,4] we used data from the two resurveys to estimate the mean usual amount (in portions per month) of intake for each baseline consumption frequency group (Supplementary Appendix S2). The group mean levels of usual consumption were used to plot against the HRs and to yield an estimate of the effect size per 1 daily portion (i.e., 100 gram/day) of fruit consumption through the Cox regression analyses.

Additional analyses were also performed to investigate the potential impacts of other dietary factors on the observed associations between fresh fruit and cardiovascular disease. Sensitivity analyses were also performed excluding the first 2 years of follow-up. All analyses were conducted in SAS (version 9.2), and graphs were plotted in R 3.0.2.

**Table S2. Repeatability of fresh fruit consumption assessed by simple qualitative food frequency questionnaire on two occasions**

| Baseline | Repeated survey (n = 926) | | | | | |
|---|---|---|---|---|---|---|
| | Never/rarely | Monthly | 1-3 days/week | 4-6 days/week | Daily | **Total** |
| Never/rarely | 11 (27.5%) | 13 | 10 | 1 | 5 | **40 (4.3%)** |
| Monthly | 13 | 77 (43.0%) | 66 | 9 | 14 | **179 (19.3%)** |
| 1-3 days/week | 9 | 37 | 130 (50.6%) | 39 | 42 | **257 (27.8%)** |
| 4-6 days/week | 4 | 17 | 27 | 23 (21.9%) | 34 | **105 (11.4%)** |
| Daily | 5 | 19 | 56 | 38 | 227 (65.8%) | **345 (37.2%)** |
| **Total** | **42 (4.5%)** | **163 (17.6%)** | **289 (31.2%)** | **110 (11.9%)** | **322 (34.8%)** | **926** |

Weighted kappa: 0.45; Spearman correlation coefficient: 0.55

**Table S3. Adjusted hazard ratios (HRs) of cardiovascular mortality and incidence of major cardiovascular diseases for daily versus never/rarely fresh fruit consumption, with different adjustment for other dietary factors**

| | Cardiovascular mortality | | Major coronary events | | Ischemic stroke | | Hemorrhagic stroke | |
|---|---|---|---|---|---|---|---|---|
| | HR | 95% CI¶ | HR | 95% CI¶ | HR | 95% CI¶ | HR | 95% CI¶ |
| Basic adjustment* | 0.60 | 0.52-0.69 | 0.66 | 0.55-0.78 | 0.75 | 0.70-0.81 | 0.64 | 0.54-0.77 |
| + fresh vegetables | 0.60 | 0.52-0.69 | 0.66 | 0.55-0.79 | 0.76 | 0.70-0.81 | 0.64 | 0.54-0.77 |
| + rice | 0.60 | 0.53-0.69 | 0.67 | 0.56-0.79 | 0.75 | 0.70-0.81 | 0.65 | 0.54-0.77 |
| + wheat | 0.60 | 0.52-0.69 | 0.66 | 0.55-0.79 | 0.75 | 0.70-0.81 | 0.64 | 0.54-0.77 |
| + other staple foods | 0.61 | 0.53-0.69 | 0.66 | 0.56-0.79 | 0.77 | 0.71-0.82 | 0.64 | 0.54-0.77 |
| + poultry | 0.61 | 0.54-0.70 | 0.67 | 0.56-0.80 | 0.76 | 0.71-0.82 | 0.65 | 0.57-0.78 |
| + fish | 0.62 | 0.54-0.71 | 0.67 | 0.57-0.80 | 0.76 | 0.71-0.82 | 0.67 | 0.56-0.79 |
| + eggs | 0.61 | 0.53-0.70 | 0.66 | 0.56-0.79 | 0.76 | 0.71-0.82 | 0.66 | 0.55-0.79 |
| + soybean | 0.60 | 0.52-0.69 | 0.66 | 0.56-0.79 | 0.75 | 0.70-0.81 | 0.64 | 0.54-0.76 |
| + tea | 0.60 | 0.53-0.69 | 0.66 | 0.56-0.79 | 0.76 | 0.70-0.81 | 0.65 | 0.54-0.77 |

¶ 95% CIs are un-floated.

* Variables for basic adjustment included age-at-risk, sex, region (as strata variables), education, income, smoking, alcohol consumption, physical activity, survey season and the intake of meat, dairy products and preserved vegetables. All the additionally adjusted dietary variables are in 3 categories, except for fresh vegetable which is in 2 categories.

**Table S4. Agreement between consumption frequency and quantitative categories integrating daily portion of consumption among those participated in the second resurvey (n = 25,069)**

| Frequency only | Fruit consumption category (integrating both frequency and number of portion) | | | | | | Mean portions* per day |
|---|---|---|---|---|---|---|---|
| | Never/rarely | Q1† | Q2† | Q3† | Q4† | **Total** | |
| Never/rarely | 2302 (100%) | 0 | 0 | 0 | 0 | **2302 (9.2%)** | 0 |
| Monthly | 0 | 3194 (96.3%) | 121 | 1 | 0 | **3316 (13.2%)** | 1.37 |
| 1-3 days/week | 0 | 0 | 7706 (94.5%) | 192 | 257 | **8155 (32.5%)** | 1.43 |
| 4-6 days/week | 0 | 0 | 0 | 1900 (70.8%) | 784 | **2684 (10.7%)** | 1.40 |
| Daily | 0 | 0 | 0 | 0 | 8612 (100%) | **8612 (34.4%)** | 1.68 |
| **Total** | **2302 (9.2%)** | **3194 (12.8%)** | **7827 (31.2%)** | **2093 (8.4%)** | **9653 (38.5%)** | **25069 (100%)** | **1.51** |
| Weighted kappa | 0.96 | | | | | | |

*1 portion = 100 gram.

†Except for those who never or rarely consumed fresh fruit, all other participants were separated into four groups according to the quartiles of the monthly consumption, which equals the number of days in a month consuming fresh fruit multiplied by the daily portion consumed.

**Figure S1. Percentage of participants who reported consuming fresh fruit on most days by age.**

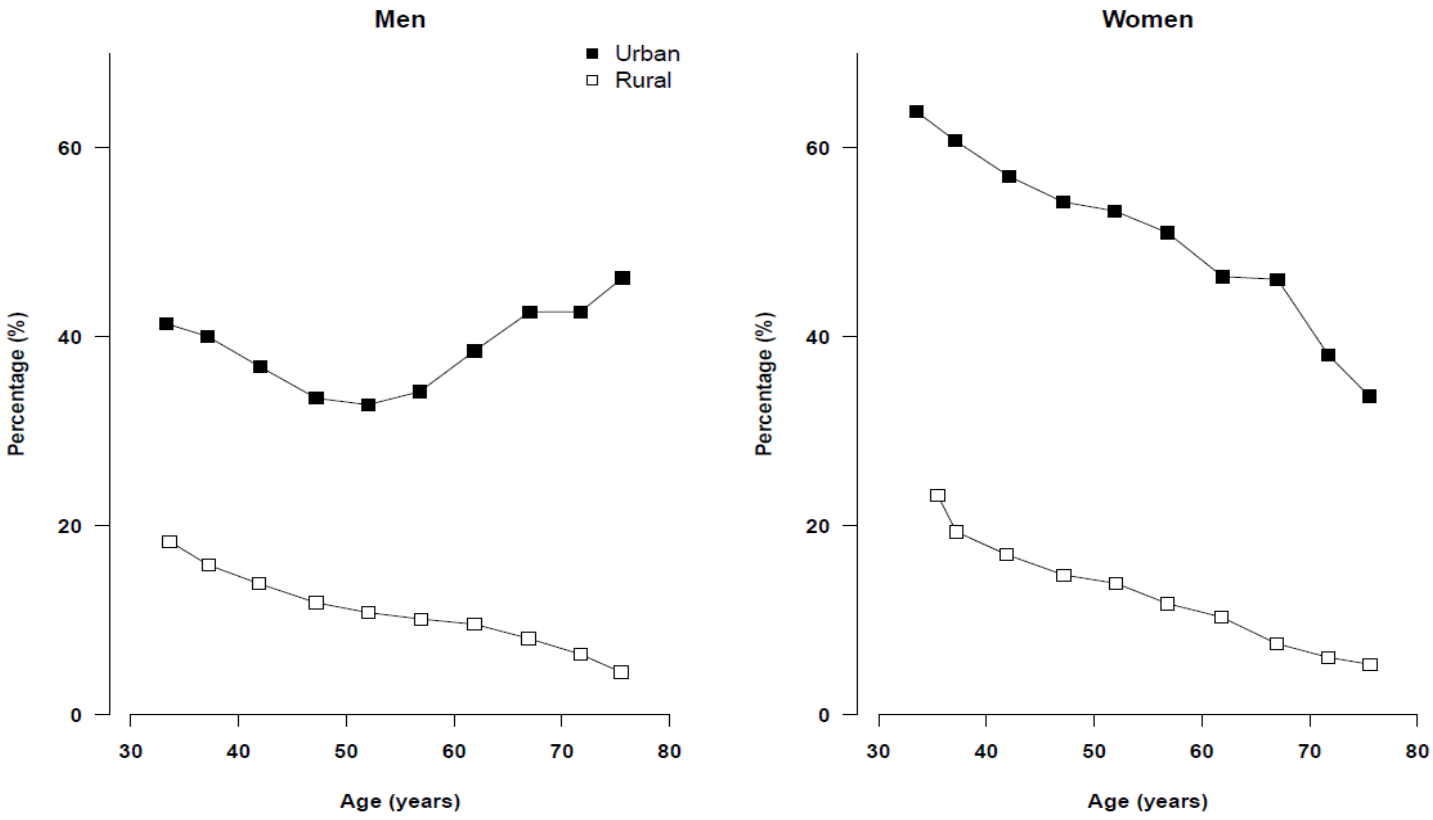


**Figure S2. Percentage of participants who reported consuming fresh fruit on most days by age and survey area.**

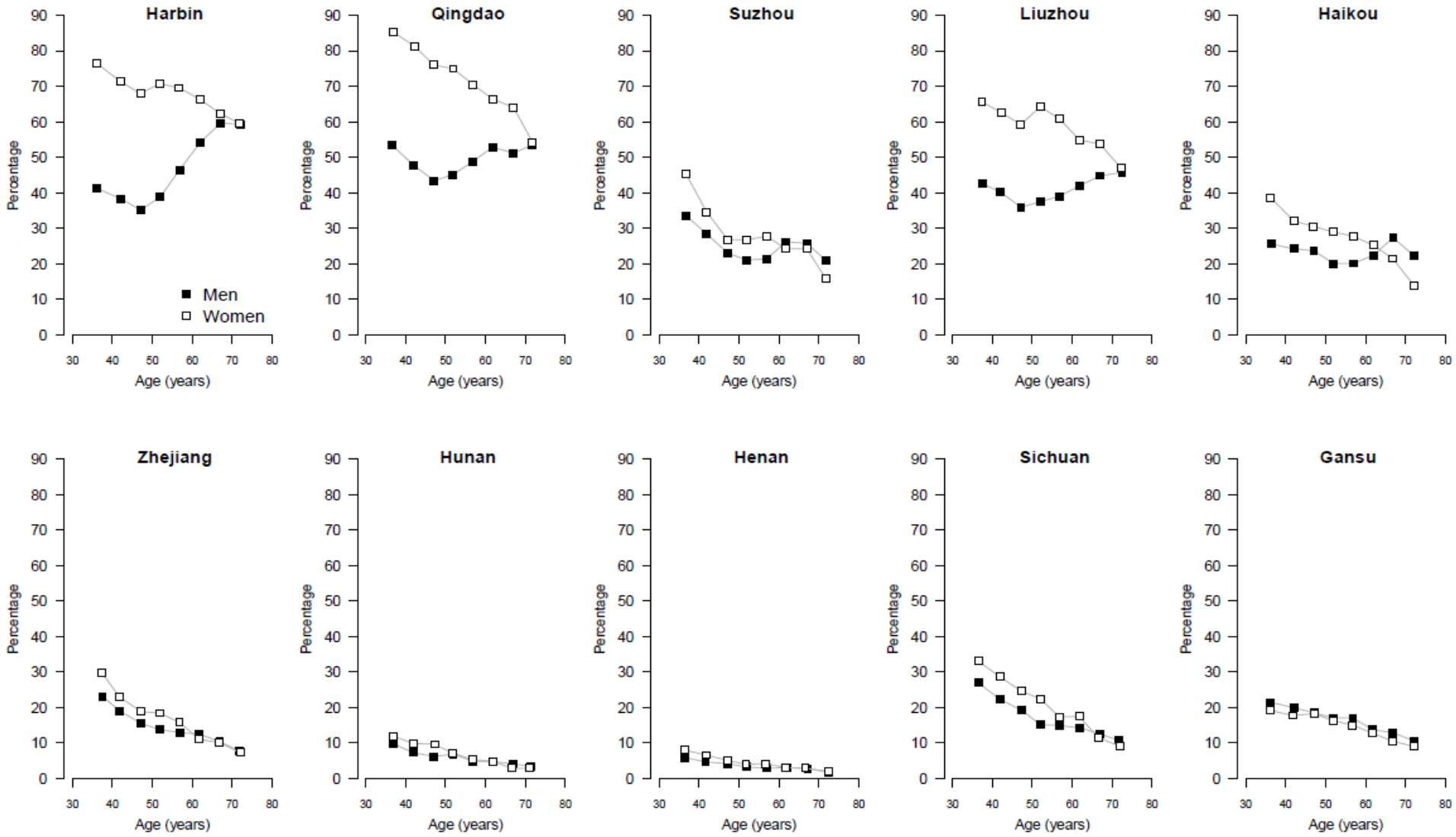
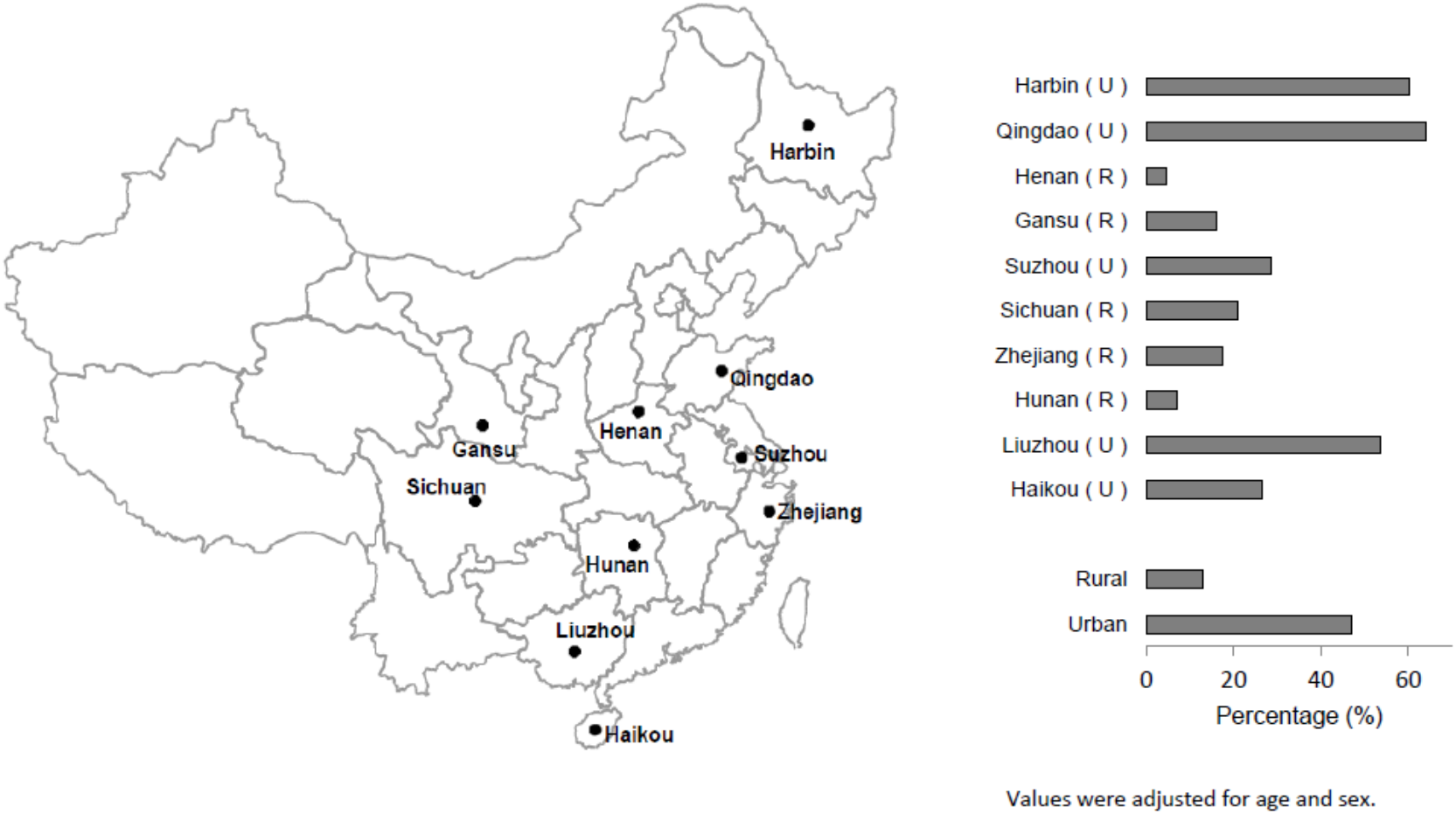
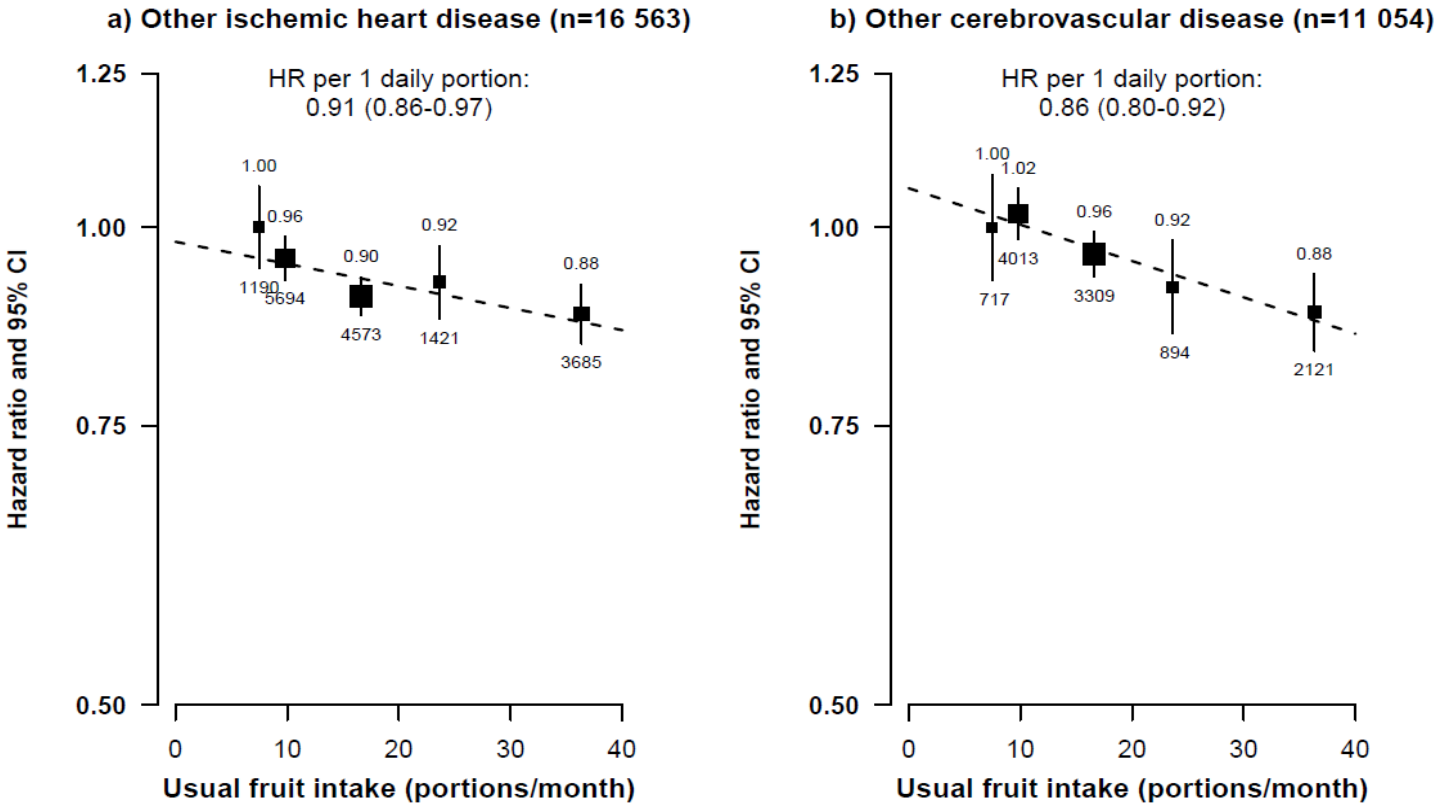


**Figure S3. Percentage of participants consuming fresh fruit on most days by area.**

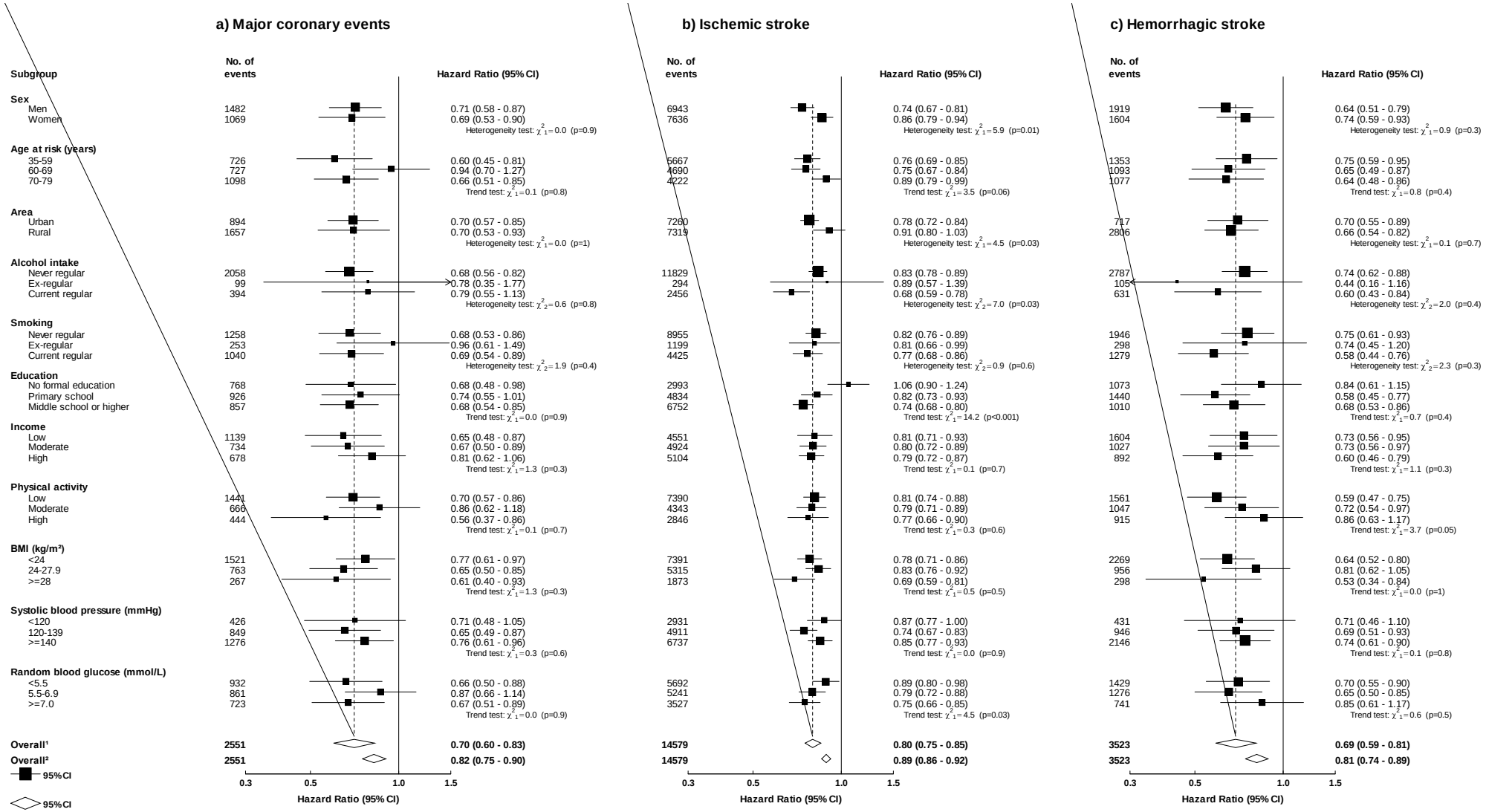



Values were adjusted for age and sex.

**Figure S4. Adjusted hazard ratios (HRs) of other cardiovascular diseases by usual habitual levels of fresh fruit consumption.**

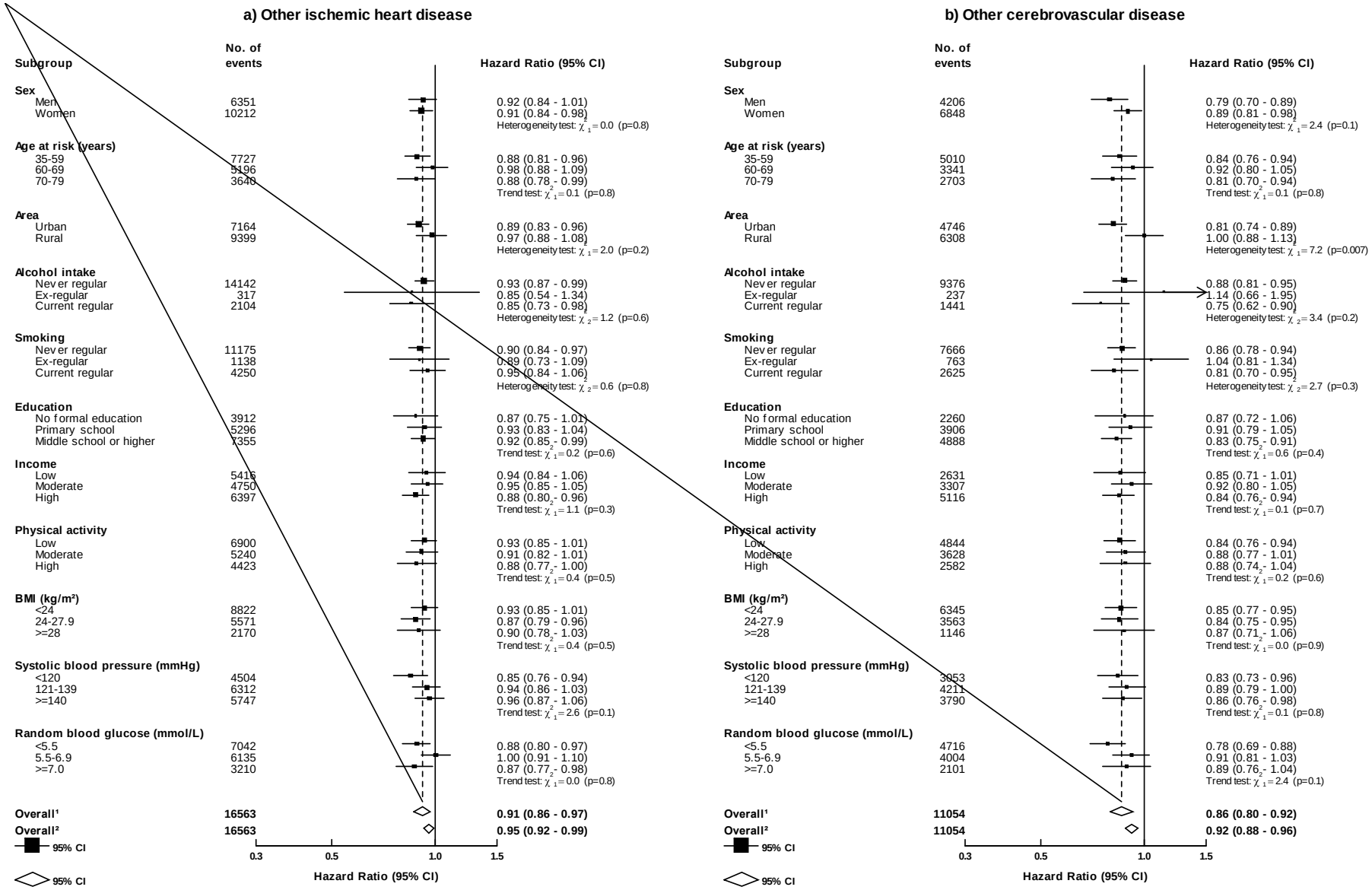


## Figure S5. Adjusted hazard ratios (HRs) of major cardiovascular diseases per 1 daily portion of fresh fruit consumption by baseline characteristics.




1. Overall effect of usual fresh fruit consumption after adjusting for regression dilution bias; 2. Overall effect of baseline fresh fruit consumption before adjusting for regression dilution bias.7503 participants with missing value on random blood glucose.

## Figure S6. Adjusted hazard ratios (HRs) of other cardiovascular diseases per 1 daily portion of fresh fruit consumption by baseline characteristics.




1. Overall effect of usual fresh fruit consumption after adjusting for regression dilution bias; 2. Overall effect of baseline fresh fruit consumption before adjusting for regression dilution bias. 7503 participants with missing value on random blood glucose.

**Figure S7. Adjusted hazard ratios (HRs) of major cardiovascular diseases per 1 daily portion of fresh fruit consumption by region.**

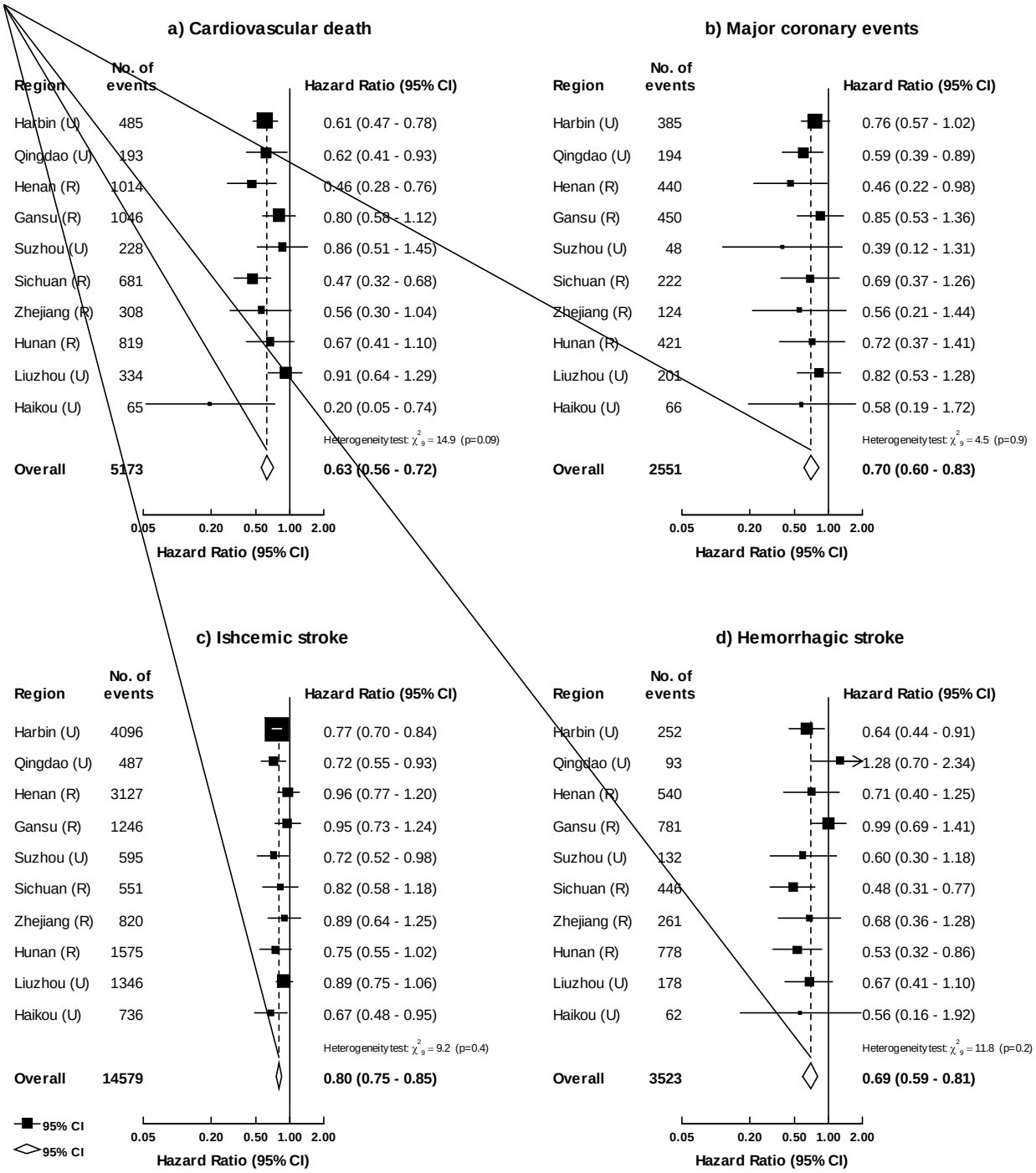


**Figure S8. Adjusted hazard ratios (HRs) of major cardiovascular diseases per 1 daily portion of fresh fruit consumption by season of enrolment.**

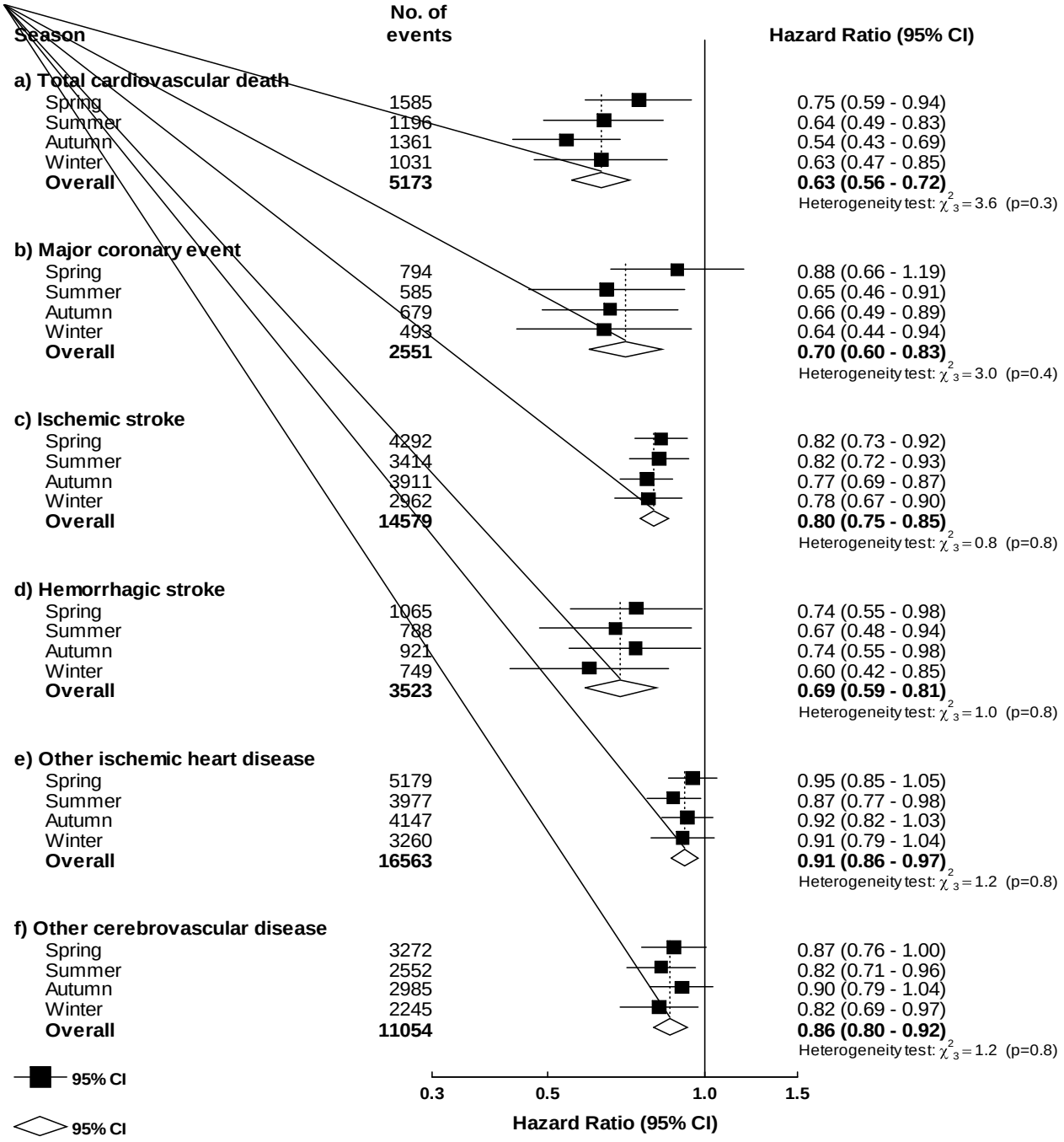
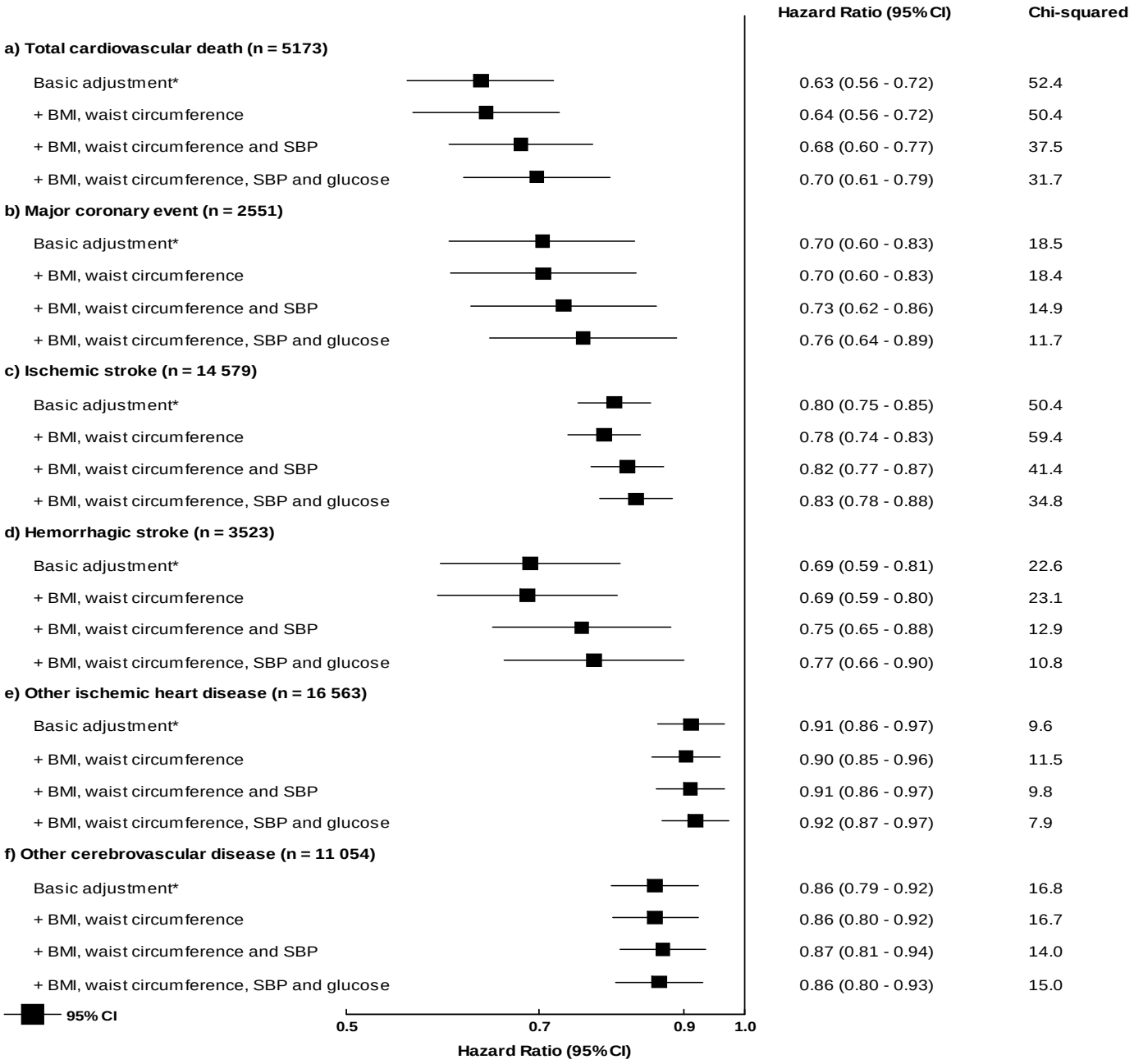


**Figure S9. Adjusted hazard ratios (HRs) of major cardiovascular events per 1 daily portion of fresh fruit consumption with different degree of adjustment for potential mediators.**




The likelihood ratio chi-squared values indicate the strength of the associations of fresh fruit consumption with disease outcomes. A larger chi-squared value indicates a stronger association and a decrease in the chi-squared value indicates that the association is attenuated after additional adjustment for the newly added variable.
*Adjustment for age, sex, region, education, income, smoking, alcohol consumption, survey season, physical activity and the consumption of meat, dairy products and preserved vegetables.

## Supplementary References


1. Plummer M. Improved estimates of floating absolute risk. Stat Med 2004;23:93-104.

2. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health 1998;88:15-9.

3. Clarke R, Shipley M, Lewington S, et al. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. Am J Epidemiol 1999;150:341-53.

4. Lewington S, Thomsen T, Davidsen M, Sherliker P, Clarke R. Regression dilution bias in blood total and high-density lipoprotein cholesterol and blood pressure in the Glostrup and Framingham prospective studies. J Cardiovasc Risk 2003;10:143-8.