

Supplementary data

The inverse association of body mass index with lung cancer: exploring residual confounding, metabolic aberrations and within-person variability in smoking

Angela M Wood, Håkan Jonsson, Gabriele Nagel, Christel Häggström, Jonas Manjer, Hanno Ulmer, Anders Engeland, Emanuel Zitt, Sylvia HJ Jochems, Sara Ghaderi, Pär Stattin, Tone Bjørge, Tanja Stocks

Supplementary Table 1. International Classification of Diseases (ICD) codes used for categorisation of cancers by topography

Supplementary Table 2. Population and analysis used to study the influence of i) metabolic aberrations, ii) residual confounding and iii) within-person variability in smoking on the associations

Supplementary Table 3. Baseline characteristics of all individuals and restricted to individuals with repeat measurements on smoking

Supplementary Figure 1. Study-specific and overall hazard ratio of lung cancer for smokers and ex-smokers *vs* non-smokers

Supplementary Figure 2. Overall and cohort and time-specific regression dilution ratios of body mass index and smoking

Supplementary Figure 3. Hazard ratio of cancer by baseline body mass index level using fractional polynomials

Supplementary Figure 4. Hazard ratio of cancers of the oral cavity, pharynx, larynx, and oesophagus-squamous cell carcinoma combined across quintiles of baseline height in subgroups of smoking, excluding the Vorarlberg Health Monitoring and Prevention Programme

Supplementary Figure 5. Hazard ratio of cancer per standard deviation of usual body mass index level among women and men, respectively

Supplementary Figure 6. Hazard ratio of lung cancer by level of body mass index and metabolic score combined, by sex, lung cancer type, smoking status and excluding the first years of follow-up

Supplementary Table 1. International Classification of Diseases (ICD) codes used for categorisation of cancers by topography

Cancer form	ICD-7	ICD-10
Liver	155.0	C22
Pancreas	157	C25
Oral cavity	140, 141.7-9, 143, 144	C00, C02-06
Pharynx	141.0, 145, 147	C01, C09-10, C12-13
Oesophagus	150	C15
Stomach	151	C16
Cardia	151.1	C16.0
Non-cardia	151.0, 151.8-9	C16.1-9
Larynx	161	C32
Lung (including bronchus)	162.1	C34
Cervix uteri	171	C53
Kidney	180	C64-65
Renal cell carcinoma	180.0, 180.9	C64
Pelvis	180.1	C65
Bladder	181.0, 181.6	C67, D09.0
Myeloid leukemia	205	C92, D46

Supplementary Table 2. Population and analysis used to study the influence of i) metabolic aberrations, ii) residual confounding and iii) within-person variability in smoking on the associations between BMI and lung cancer, and other smoking-related cancers

Influence of	Population	Analysis
Metabolic aberrations	488,659 individuals with complete information on MS (blood pressure, triglycerides and glucose)	Combined association of categories of BMI (<25/≥25 kg/m ²) and MS (</≥ median value) with cancer risk
Residual confounding by smoking	227,198 non-smokers, excluding from the VHM&PP, and 207,630 smokers	Association between BMI and cancer risk amongst non-smokers vs smokers
Within-person variability in smoking	All 778,828 individuals in the study	Association between BMI and cancer risk after adjustment for within-person variability in smoking (“usual smoking”), calculated from 221,958 individuals with 600,201 serial assessments of smoking status and pack-years

MS, metabolic score; BMI, body mass index.

Supplementary Table 3. Baseline characteristics of participants in the Metabolic syndrome and Cancer project 2.0, with and without repeat data on smoking

Characteristic	All individuals (n=778 828)	Individuals with repeat smoking information (n=221 958)
Cohort, n (%)		
VHM&PP	161 232 (21)	118 398 (53)
VIP	94 205 (12)	31 958 (14)
MPP	20 971 (3)	1396 (1)
40-y	394 722 (51)	6719 (3)
Oslo	17 039 (2)	254 (0.1)
NCS	90 659 (12)	63 233 (28)
Age, mean (SD) years	44 (9)	41 (12)
Baseline smoking status, n (%)		
Non-smoker	357 147 (46)	138 286 (60)
Ex-smoker	182 858 (23)	28 762 (12)
Smoker, <20 pack-years	181 245 (23)	36 516 (16)
Smoker, ≥20 pack-years	57 579 (8)	15 457 (7)
Smoker, missing packyears	0 (0)	11 548 (5)
Male, n (%)	382 853 (49)	108 546 (48)
Decade of birth, n (%)		
1890-1930	62 034 (8)	28 473 (12)
1930-1950	299 957 (39)	106 685 (46)
1950-1970	383 530 (49)	85 494 (37)
1970-1990	33 307 (4)	9917 (4)
BMI, mean (SD) kg/m ²	25.0 (3.9)	25.7 (4.2)
BMI category, kg/m ² , n (%)		
<20	47 069 (6)	12 995 (6)
20-22.49	200 814 (26)	49 245 (21)
22.5-24.99	181 630 (23)	48 549 (21)
25-27.49	178 351 (23)	54 315 (24)
27.5-29.99	92 765 (12)	32 398 (14)
≥30	78 199 (10)	33 067 (14)
Metabolic score, mean (SD)	-0.05 (1.99)	-0.05 (1.92)
BMI-metabolic score category ^a , n (%)		
BMI<25 kg/m ² -high metabolic score	93 941 (19)	43 042 (21)
BMI<25 kg/m ² -low metabolic score	165 776 (34)	74 158 (36)
BMI≥25 kg/m ² -high metabolic score	147 523 (31)	60 294 (29)
BMI≥25 kg/m ² -low metabolic score	77 977 (16)	28 712 (14)
Mid-blood pressure, mean (SD) mmHg ^b	105 (13)	107 (14)
Serum triglycerides, mean (SD) mmol/l ^{b,c}	1.4 (1.0)	1.4 (1.1)
Serum glucose, mean (SD) mmol/l ^{b,d}	5.4 (1.2)	5.4 (1.3)
Diabetes ^e , n (%)	6007 (1)	1728 (1.5)

Oslo, Oslo study I; NCS, Norwegian Counties Study; 40-y, Age 40-Programme; VHM&PP, Vorarlberg Health Monitoring and Prevention Programme; VIP, Västerbotten Intervention Project; MPP, Malmö Preventive Project; SD, standard deviation; BMI, body mass index.

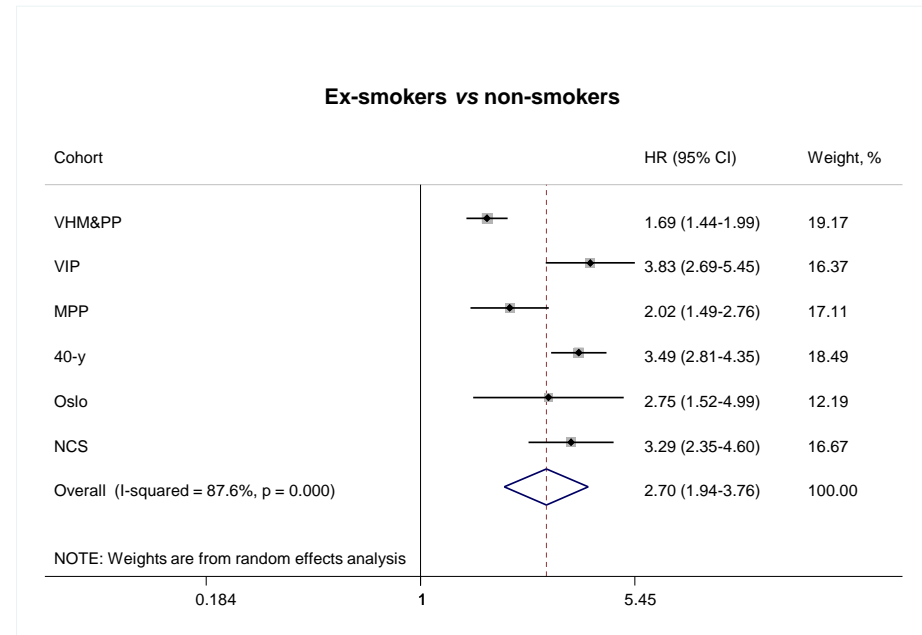
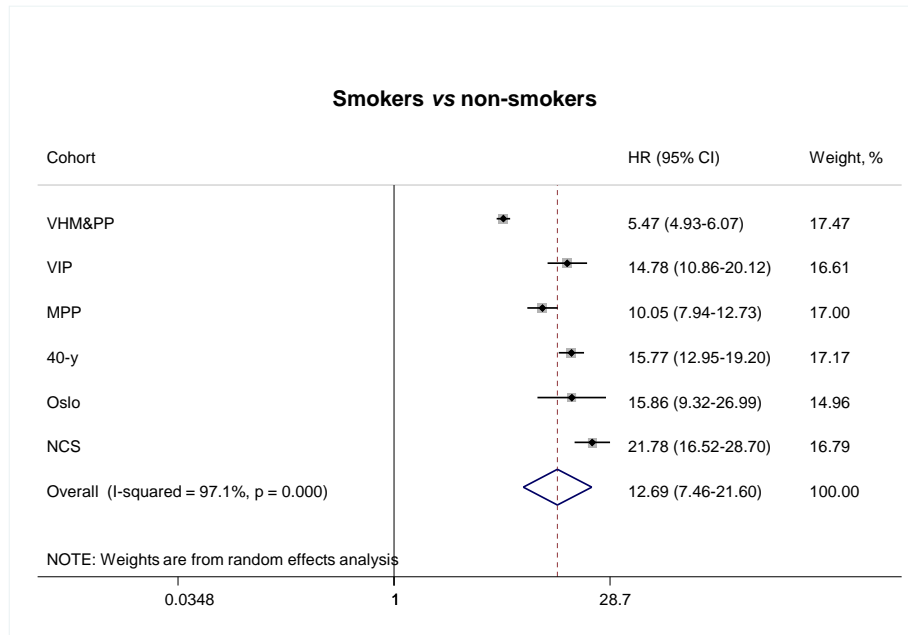
^aThe metabolic score is a sum of z scores for mid-blood pressure, triglycerides and glucose, each standardised by cohort, sex, and fasting time (except mid-blood pressure), calculated for 485 217 (62%) individuals in the full sample and 206 206 (93%) individuals with repeat data. The median was used as cut-point for low/high metabolic score.

^bMean and SD calculated separately by cohort and then combined using random effects meta-analysis.

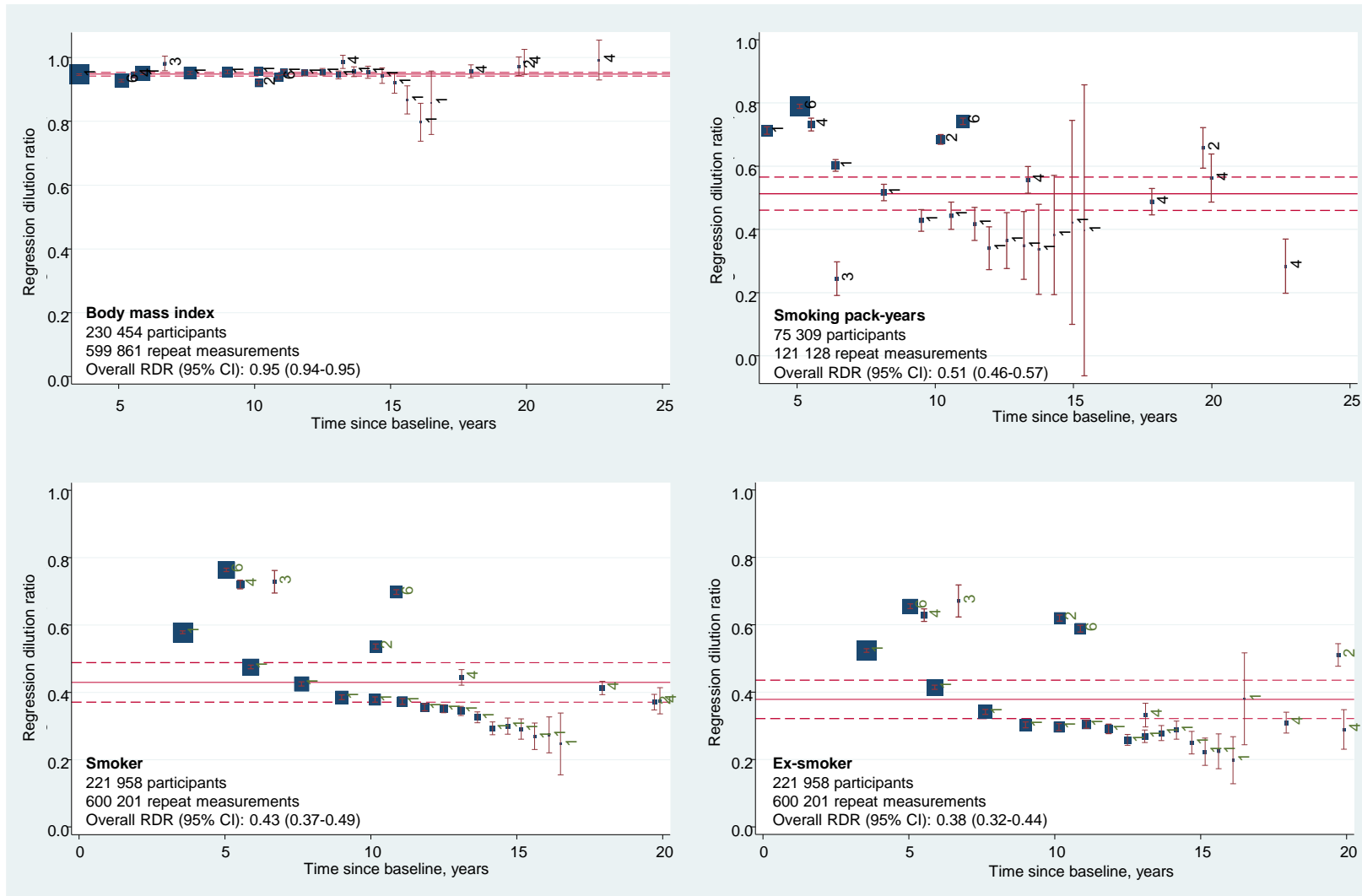
^cBased on 286 249 (37%) and 152 999 (66%) individuals with eight hours or more of fasting.

^dBased on 260 038 (33%) and 150 700 (65%) individuals with eight hours or more of fasting and with glucose measured in serum or plasma (excludes the MPP where glucose was measured in whole blood).

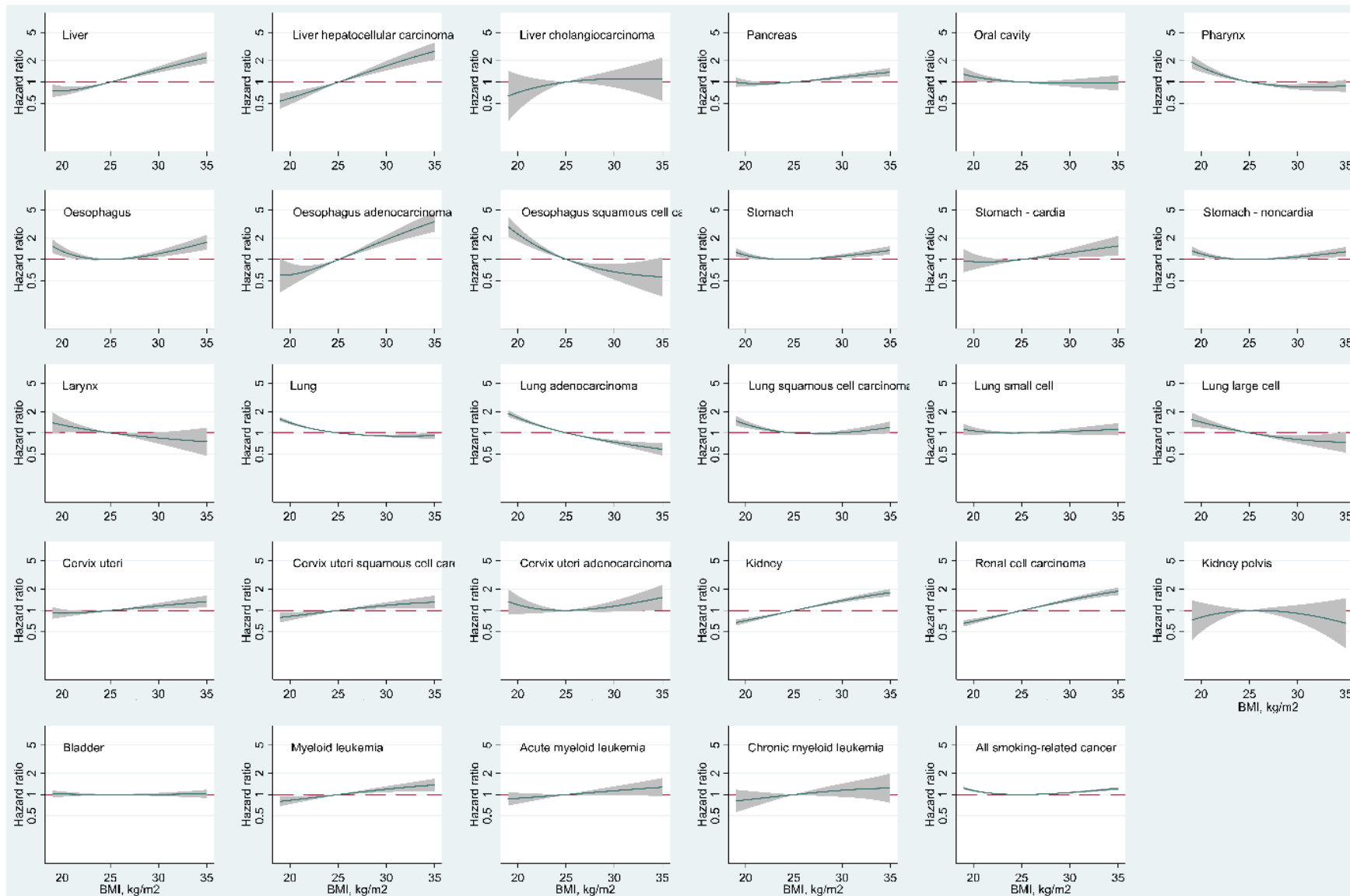
^eSelf-reported diabetes. Excludes the VHM&PP that lacked the information.



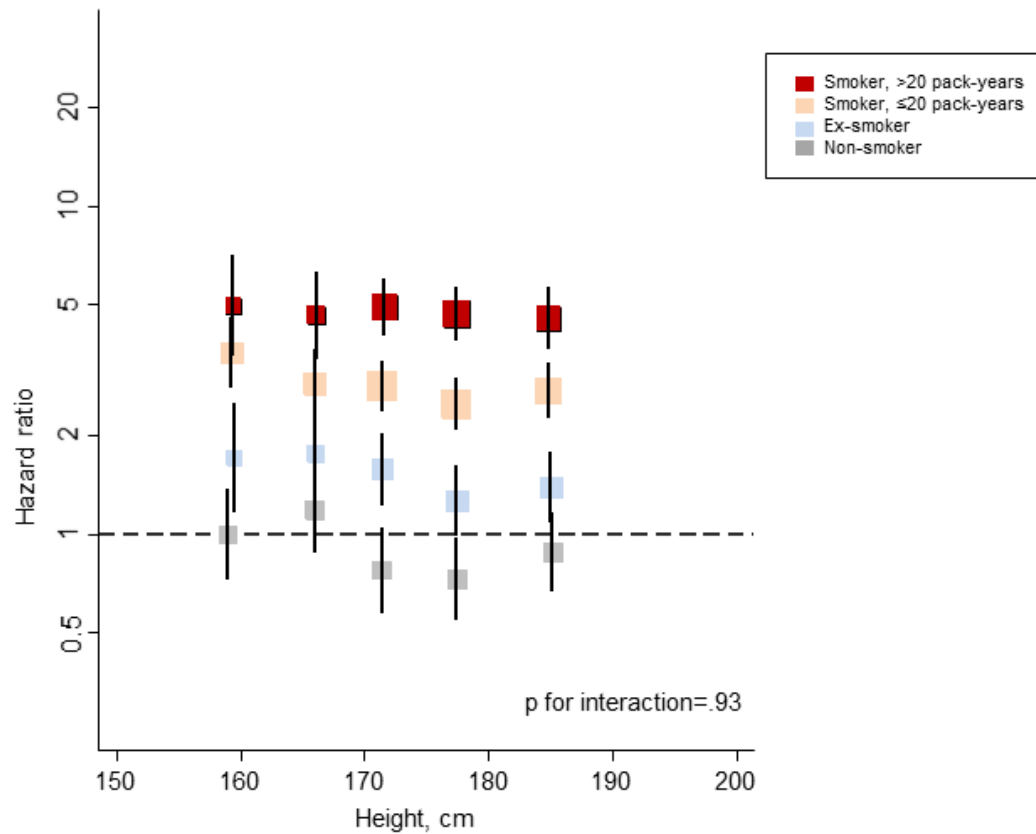
Supplementary Figure 1. Study-specific and overall hazard ratio (95% confidence interval) of lung cancer for smokers (n=238 823, cases=5090) and ex-smokers (n=201 018, cases=865) vs non-smokers (n=357 147, cases=1054). Abbreviations: VHM&PP, Vorarlberg Health Monitoring and Prevention Programme; VIP, Västerbotten Intervention Project; MPP, Malmö Preventive Project; 40-y, Age 40-programme; Oslo, Oslo study I; NCS, Norwegian Counties Study.



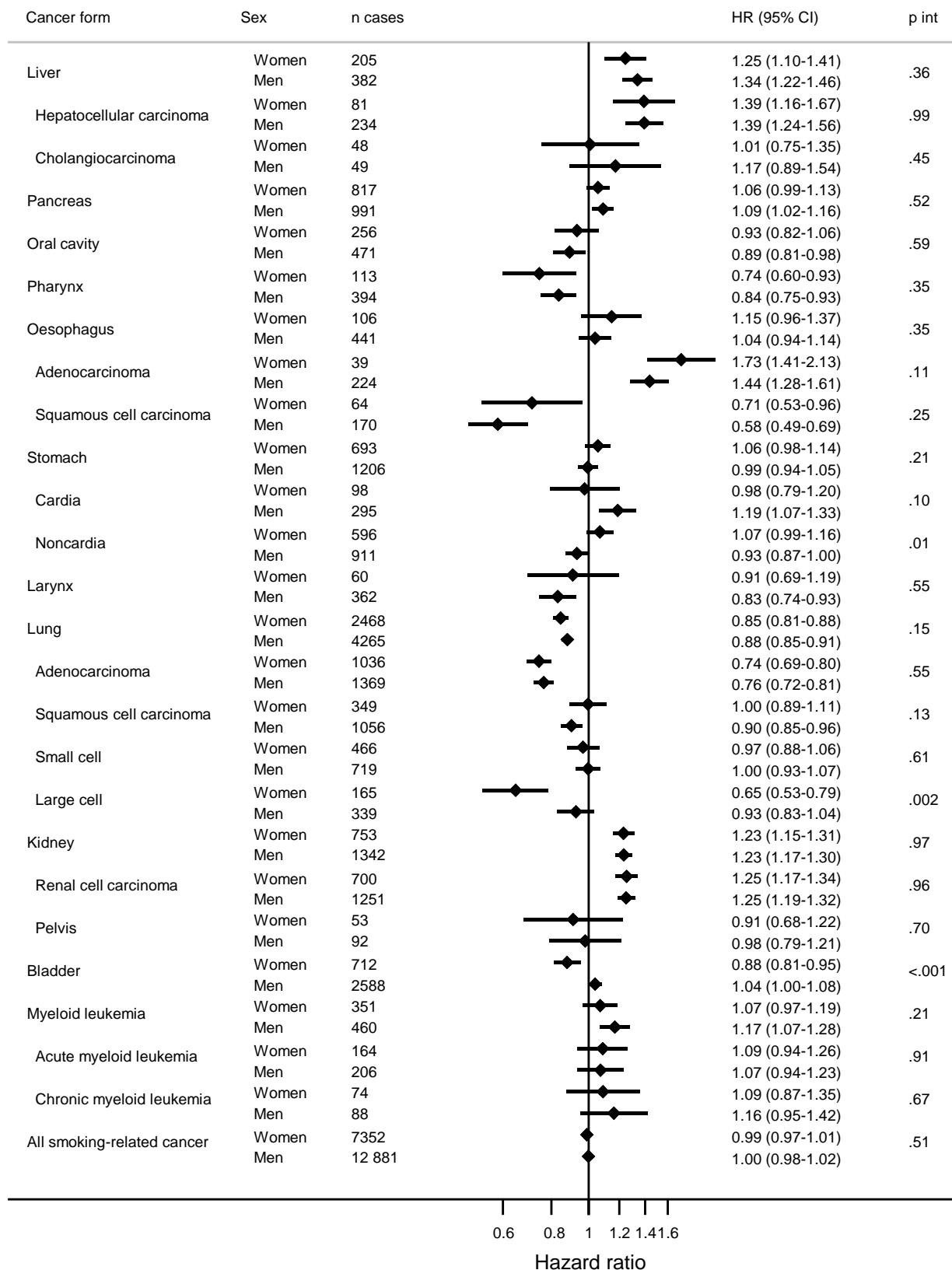
Supplementary Figure 2. Overall and cohort and time-specific regression dilution ratios (95% confidence interval) of body mass index and smoking. Cohort enumeration: 1=Vorarlberg Health Monitoring and Prevention Programme, 2=Västerbotten Intervention Project, 3=Malmö Preventive Project, 4=Age 40-Programme, 5=Oslo Study I, 6=Norwegian Counties Study.



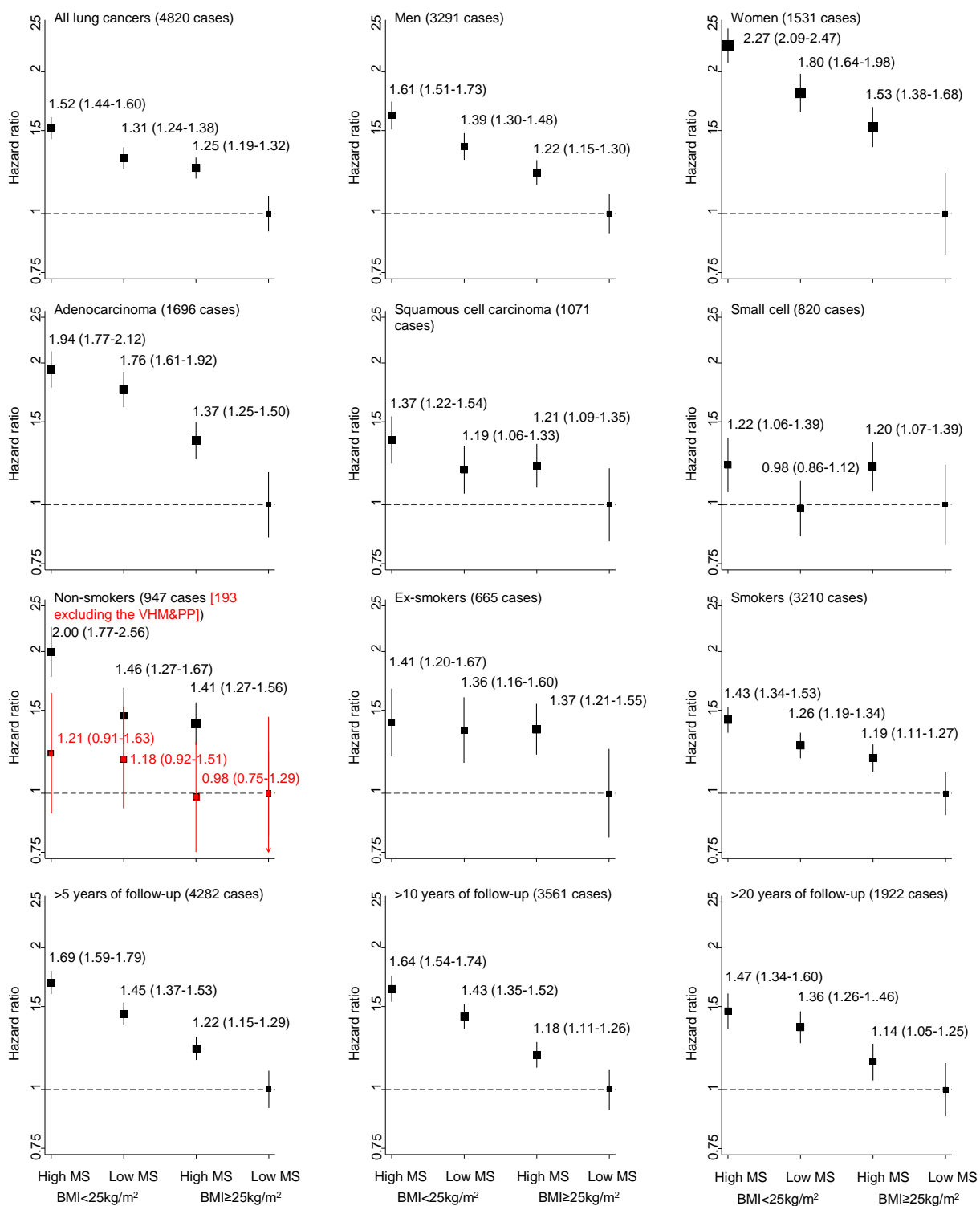
Supplementary Figure 3. Hazard ratio (95% confidence interval) of cancer by baseline body mass index level (n=778 828). Hazard ratios were calculated using the best fitting first or second order fractional polynomial using Cox regression with attained age as time-scale, stratified by cohort, sex, and birth decade, and adjusted for baseline age, fasting status, smoking status and pack-years.



Supplementary Figure 4. Hazard ratio (95% confidence interval) of cancers of the oral cavity, pharynx, larynx, and oesophagus-squamous cell combined (n cases=1392) across quintiles of baseline height in subgroups of smoking (n participants=617 596), excluding the Vorarlberg Health Monitoring and Prevention Programme. Hazard ratios were calculated using Cox regression with attained age as time-scale, stratified by cohort, sex, and birth decade. Non-smokers in the lowest height quintile is the referent group. P for interaction denotes the p-value for the difference in linear trends of hazard ratios across levels of height between smoking groups.



Supplementary Figure 5. Hazard ratio (95% confidence interval) of cancer per standard deviation of usual body mass index level among women (n=395 975) and men (n=382 853), respectively. Hazard ratios were calculated using Cox regression with attained age as time-scale, stratified by cohort, sex, and birth decade, and adjusted for baseline age, fasting status, usual smoking status, and usual pack-years. P int denotes the p-value for interaction between body mass index and sex.



Supplementary Figure 6. Hazard ratio (95% confidence interval) of lung cancer by level of body mass index and metabolic score combined (n participants=503 167), by sex, lung cancer type, smoking status and excluding the first five, 10 and 20 years of follow-up. The metabolic score comprises mid-blood pressure, serum triglycerides and glucose, each standardised by cohort, sex, and fasting time (except mid-blood pressure) before being summed up and divided at the median (high/low MS). Hazard ratios were calculated using Cox regression with attained age as time-scale, stratified by cohort, sex, and birth decade, and adjusted for baseline age, usual smoking status, and usual pack-years. BMI ≥ 25 kg/m² with low MS is the referent group.